BLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

D. DUGUÉ et M. GIRAULT

FONCTIONS CONVEXES DE POLYA

M. FRÉCHET

COURBES ALÉATOIRES

E. HALPHEN

LES FONCTIONS FACTORIELLES

E. HALPHEN

LA NOTION DE VRAISEMBLANCE

J.-L. SOULÉ

ÉNERGIE CINÉTIQUE MOYENNE D'UNE PARTICULE EN MOUVEMENT BROWNIEN DANS UN FLUIDE RÉEL

VOL. IV - FASCICULE 1 - 1955

PARIS
11, Rue Pierre Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse :

INSTITUT DE STATISTIQUE DE L'UNIVERSITE DE PARIS

INSTITUT HENRI POINCARÉ - 11, Rue Pierre Curie - PARIS (Ve)

Les manuscrits doivent être envoyés à *M. Daniel DUGUÉ*,
à l'adresse précédente.

Abonnements : Pour la France 1.700 francs français
Pour l'Etranger 2.000 francs

Vente au numéro : *(fascicule de 50 pages environ)*
Pour la France 500 francs
Pour l'Etranger 600 francs

PUBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

D. DUGUÉ et M. GIRAULT

FONCTIONS CONVEXES DE POLYA

M. FRÉCHET

COURBES ALÉATOIRES

E. HALPHEN

LES FONCTIONS FACTORIELLES

E. HALPHEN

LA NOTION DE VRAISEMBLANCE

J.-L. SOULÉ

ÉNERGIE CINÉTIQUE MOYENNE D'UNE PARTICULE EN MOUVEMENT BROWNIEN DANS UN FLUIDE RÉEL

VOL. IV - FASCICULE 1 - 1955

PARIS
11, Rue Pierre Curie



FONCTIONS CONVEXES DE POLYA

par

D. DUGUÉ & M· GIRAULT

Polya a donné des conditions suffisantes simples pour qu'une fonction φ (t) soit une fonction caractéristique (f.c.). (1)

Ces conditions sont les suivantes :

(cond. I)
- φ (t) réelle, définie continue pour toutes les valeurs réelles de t
- $\varphi(0) = I \qquad \varphi(-t) = \varphi(t)$
- lim pour $t = \infty$ de $\varphi(t) = 0$
- φ (t) convexe pour $t > 0$

C'est essentiellement cette dernière condition qui fait l'originalité de la catégorie des fonctions et signifie que la courbe représentant φ (t) a sa concavité vers le haut ($\Delta \varphi'_t \geqslant 0$ pour $t > 0$). Les f.c. ainsi définies sont dites fonctions convexes.

DÉMONSTRATION GÉOMÉTRIQUE DU THÉORÈME DE POLYA

Remarquons tout d'abord que si φ (t), fonction convexe est f.c., elle définit une répartition continue admettant une densité f(x) puisque $\varphi(\infty) = 0$ Cette densité est donnée par la formule de Fourier.

$$f(x) = \lim_{c=\infty} \frac{1}{2\pi} \int_{-c}^{+c} e^{-itx} \varphi(t) \, dt$$

soit en tenant compte de la parité de φ

$$(2) = \lim_{c=\infty} \frac{1}{\pi} \int_{0}^{c} \varphi(t) \cos tx \, dt$$

Il reste à montrer que cette densité n'est jamais négative. La formule (2) montre que f(x) est l'abscisse, dans le plan complexe (et à un

---

(1) Remarks on characteristic functions - Proccedings of the Berkeley Symposium.

coefficient positif près) du barycentre de "masses" enroulées sur le cercle trigonométrique. La densité curviligne de ces masses est $\frac{1}{x}\varphi\left(\frac{\alpha}{x}\right)$ au point de coordonnée angulaire $\alpha = tx$.

X étant donné, décomposons l'intervalle de variation de t en éléments qui correspondent à 1 tour sur le cercle trigonométrique à partir de l'origine. Les intervalles partiels sont limités par les valeurs :

$$0,\ \frac{2\pi}{x},\ 2\,\frac{2\pi}{x}\ \ldots,\ \frac{N\cdot 2\pi}{x};\ (N+1)\,\frac{2\pi}{x}\ldots$$

Si t augmente de $N.\frac{2\pi}{x}$ à $(N+1)\frac{2\pi}{x}$ le point correspondant, d'abscisse angulaire tx fait 1 tour exactement et la densité décroit régulièrement comme le montre la figure 1. Montrons que l'abscisse du centre de gravité partiel de

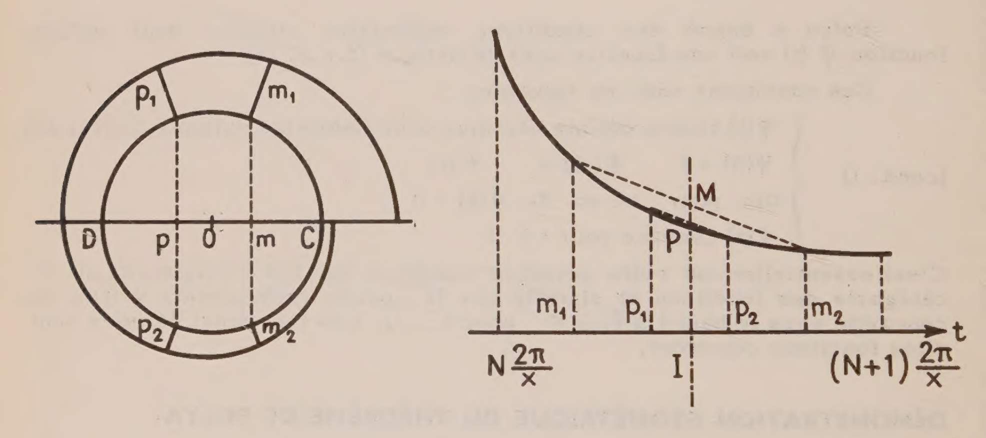


Fig. 1 **Fig. 2**

de ces masses est positif. En effet si l'on réunit ces masses sur le demi-cercle supérieur ; on obtient en 2 points d'abscisses om et op les densités respectives.

$$m_1 + m_2 = 2\,IM$$
$$p_1 + p_2 = 2\,IP$$

Si la courbe $y = \varphi(t)$ est toujours convexe du côté de l'axe des t :

$$IM \geqslant IP \quad \text{si } cm < cp$$

C'est dire que la densité sur $\widehat{CD}$ est régulièrement décroissante de C à D et par suite, le centre de gravité de ces masses est sur la partie positive OC.

Cette démonstration rend intuitif le lien entre la convexité de φ (t) et le fait que dF (x) soit positif ou nul. Elle montre aussi que la condition d'avoir un centre de gravité général d'abscisse positive, pour toute valeur de x(condition suffisante) est ici réalisée d'une manière particulièrement étroite. Les fonctions convexes sont donc une classe très spéciale de f.c.

Remarque : La condition : lim. pour $t \longrightarrow \infty$ de $\varphi(t) = 0$ est superflue. Elle peut être remplacée par la condition

$$\varphi(t) \geqslant 0$$

Désignons par (cond. II) le nouvel ensemble d'hypothèses. Si $\varphi(t)$ réalise les cond. I $\varphi_1(t) = p + q\,\varphi(t)$ est f.c. et réalise les conditions II. (p et q sont des entiers positifs de somme 1).

Réciproquement toute fonction $\varphi_1(t)$ satisfaisant aux conditions II peut se mettre sous la forme $p + q\,\varphi(t)$ où

$p = \lim_{t \to \infty} \varphi_1(t)$ est positif ou nul

$q > 0$ et $\varphi(t)$ satisfait aux conditions I

Cas particulier :

Un cas particulier très simple de fonctions convexes est constitué par des fonctions $k\left(\frac{t}{a}\right)$

$$k\left(\frac{t}{a}\right) = 1 - \left|\frac{t}{a}\right| \qquad \text{pour } |t| < a$$
$$k\left(\frac{t}{a}\right) = 0 \qquad \text{pour } |t| > a$$

Cet exemple avait été donné par Khintchine.

Les fonctions $k\left(\frac{t}{a}\right)$ jouent en quelque sorte le rôle de fonctions convexes élémentaires car à partir d'elles on peut de plusieurs manières former toutes les fonctions convexes : par mélange de lois et par produit.

MÉLANGE DE LOIS CONVEXES $k\left(\frac{t}{a}\right)$

Par mélange d'un nombre fini n de lois convexes ayant pour f.c. $k\left(\frac{t}{a}\right)$ on obtient des lois ayant pour f.c. :

$$\varphi(t) = p_1\, k\left(\frac{t}{a_1}\right) + \ldots\ldots\ldots + p_n\, k\left(\frac{t}{a_n}\right)$$
$$\text{où } \sum p_h = 1 \quad p_h \geqslant 0 \text{ et } 0 < a_1 < a_2 \ldots\ldots < a_n$$

On vérifie facilement que φ est une f.c. convexe ; c'est un contour polygonal symétrique défini par ses sommets :

$$t = a_n, \quad y = \varphi(a_n) = 0$$
$$t = a_{n-1}, \quad y = \varphi(a_{n-1}) = p_n\, k\left(\frac{a_{n-1}}{a_n}\right) = p_n\left(1 - \frac{a_{n-1}}{a_n}\right)$$
$$(3) \quad t = a_{n-2}, \quad y = \varphi(a_{n-2}) = p_{n-1}\, k\left(\frac{a_{n-2}}{a_{n-1}}\right) + p_n\, k\left(\frac{a_{n-2}}{a_n}\right)$$
$$= p_{n-1}\left(1 - \frac{a_{n-2}}{a_{n-1}}\right) + p_n\left(1 - \frac{a_{n-2}}{a_n}\right)$$

Réciproquement tout contour polygonal convexe réalisant les conditions I peut être obtenu de cette manière : les valeurs a_i sont les abscisses des sommets et leurs ordonnées déterminent les p_i par les relations (3) prises dans l'ordre indiqué.

On généralise immédiatement ce résultat au cas d'une fonction φ (t) à dérivée continue, par limite d'un mélange infini de lois convexes $k(\frac{t}{a})$. Ainsi toute fonction caractéristique convexe est la limite d'un mélange infini de lois à f.c. $k(\frac{t}{a})$.

I- Il est facile en utilisant un passage à la limite à partir du procédé précédent, de montrer qu'une loi convexe satisfaisant aux conditions 1 (c'est-à-dire telle que $\varphi(\infty) = 0$) est toujours le produit d'une variable de Khintchine par une autre variable.

En effet, soit φ (x) une fonction convexe dont la limite est nulle pour t infini. φ (x) a toujours une dérivée à gauche et une dérivée à droite. Appelons φ_G (x) la dérivée à gauche. On a en intégrant par partie :

$$\int_t^{+\infty} \left(1-\frac{t}{x}\right) d\left[1-\varphi(x)+x\,\varphi'_G(x)\right] =$$

$$\lim_{x=\infty} \left(1-\frac{t}{x}\right)\left[1-\varphi(x)+x\,\varphi'_G(x)\right] - t\int_t^{+\infty} \frac{1-\varphi(x)+x\,\varphi'_G(x)}{x^2}\,dx$$

$$= \lim_{x=\infty} \left(1-\frac{t}{x}\right)\left[1-\varphi(x)+x\,\varphi'_G(x)\right] + t\left[\frac{1}{x}\right]_t^{+\infty} - t\int_t^{+\infty}\left[\frac{d_G}{dx}\,\frac{\varphi(x)}{x}\right] dx$$

$\varphi(x) - x\,\varphi'_G(x)$ est l'ordonnée à l'origine de la tangente à gauche à la courbe $y = \varphi(x)$ au point d'abscisse x. Cette quantité est non croissante puisque φ (x) est convexe. D'autre part $\varphi(x) - x\,\varphi'_G(x)$ tend vers 0 quand x augmente indéfiniment puisque $\varphi(\infty) = 0$. De plus $\frac{d_G}{dx}\,\frac{\varphi(x)}{x}$ étant bornée on a:

$$\int_t^{+\infty} \frac{d_G}{dx}\left(\frac{\varphi(x)}{x}\right) dx = \left[\frac{\varphi(x)}{x}\right]_t^{+\infty}$$

il en résulte que :

$$\int_t^{+\infty} \left(1-\frac{t}{x}\right) d\left[1-\varphi(x)+x\,\varphi'_G(x)\right] = \varphi(t)$$

et finalement en utilisant la fonction k (t) définie plus haut :

$$\varphi(t) = \int_0^{+\infty} k\left(\frac{t}{x}\right) dF(x) \quad \text{en posant } F(x) = 1-\varphi(x)+x\,\varphi'_G(x)$$

F (x) non décroissante est telle que F (0) = 0 $F(\infty) = 1$ est une fonction de probabilité totale.

On a, puisque k (t) est fonction caractéristique :

$$k(t) = \int_{-\infty}^{+\infty} e^{ity}\,dk(y) \quad k(y) \text{ Etant une fonction de proba-}$$

bilité totale,

et par suite

$$\varphi(t) = \int_0^{+\infty} \int_{-\infty}^{+\infty} e^{it\frac{y}{x}} \, dk(y) \, dF(x)$$

φ (t) est la fonction caractéristique du quotient de 2 variables indépendantes, la variable numérateur étant une variable de Khintchine. C'est bien ce qui avait été annoncé.

Dans le même ordre d'idée, si φ (t) est une fonction caractéristique convexe et F (x) une loi de probabilité totale, la fonction

$$\Phi(t) = \int \varphi(tx) \, dF(x)$$

dont il est bien connu qu'elle est caractéristique est une fonction convexe. En effet, par définition :

$$\varphi(\alpha t_1 + \beta t_2) < \alpha \varphi(t_1) + \beta \varphi(t_2) \text{ avec } \alpha, \beta > 0 \text{ et } \alpha + \beta = 1$$

Ce qui entraine la même inégalité pour $\Phi(t)$

Donc si une variable aléatoire a pour fonction caractéristique une fonction convexe, le produit de cette variable aléatoire par une variable aléatoire indépendante quelconque a aussi pour fonction caractéristique une fonction convexe.

On peut généraliser le résultat sur la variable de Khintchine sous la forme suivante. Si φ (x) est une fonction telle que $1 - \varphi(x) + \frac{x}{\alpha} \varphi'_G(x)$ croisse de 0 à 1, quand x varie de 0 à $+\infty$ et si α est un nombre positif inférieur à 1, la variable dont la fonction caractéristique est φ (t), est le produit d'une variable dont la fonction caractéristique est $1 - |t|^{\alpha}$ pour $|t| < 1$ et zéro pour $|t| > 1$ par une autre variable indépendante.

C'est la conséquence du fait que par les hypothèses indiquées :

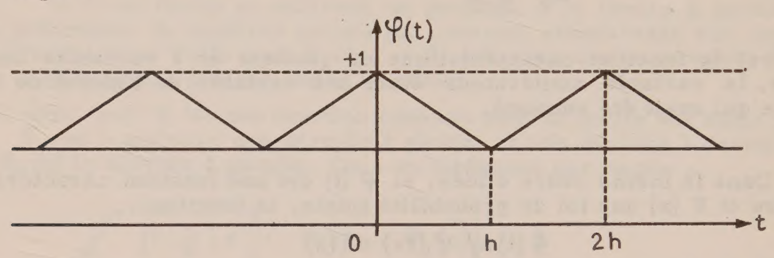
$$\varphi(t) = \int_t^{+\infty} \left[1 - \left(\frac{t}{x}\right)^{\alpha}\right] d \left[1 - \varphi(x) + \frac{x}{\alpha} \varphi'_G(x)\right]$$

et que $1 - \varphi(x) + \frac{x}{\alpha} \varphi'_G(x)$ est sur les mêmes hypothèses une loi de probabilité totale.

On voit de cette façon que toutes variables aléatoires dont la fonction caractéristique $1 - |t|^{\alpha}$ pour $|t| < 1$ et 0 pour $|t| > 1$ sont le produit d'une variable aléatoire dont la fonction caractéristique est $1 - |t|^{\beta}$ pour $|t| < 1$ et 0 pour $|t| > 1$ (α et β étant inférieurs à 1) par une autre v.a. indépendante.

II- Les variables aléatoires ayant pour loi de probabilité des fonctions convexes donnent lieu à ce que nous proposons d'appeler le phénomène de Khintchine (voir par exemple P. Lévy-Théorie d'addition des variables aléatoires p. 194 8°) et qui est le fait que deux variables aléatoires peuvent avoir des fonctions caractéristiques coincidant sur un segment entourant l'origine sans coincider sur tout l'axe réel. C'est à peu près évident, car on peut très bien imaginer 2 fonctions convexes coincidant sur un segment entourant l'origine et distinctes hors de ce segment.

Nous allons établir maintenant le résultat suivant : une fonction $\varphi(t)$ égale à 1 pour $t = 0$, réelle et symétrique, qui dans l'intervalle (oh) est convexe et qui est périodique de période 2 h (voir figure) est une fonction caractéristique.



Il suffira d'établir, si l'on appelle a l'ordonnée minimum de $\varphi(t)$ que $\varphi_1(t) = (1/1-a)(\varphi(t)-a)$ est une fonction caractéristique, car alors $(1-a)\,\varphi_1(t) + a$ le sera aussi. Considérons la fonction $\varphi_2(t)$ qui coincide avec $\varphi_1(t)$ entre $-h$ et $+h$ et qui est nulle en dehors de cet intervalle. Elle est convexe donc caractéristique d'après le théorème de Polya. On a :

$$\lim_{c=\infty} \frac{1}{2\pi}\int_{-c}^{c} e^{itx}\,\varphi_2(t)\,dt = \frac{1}{2\pi}\int_{-h}^{+h} e^{-itx}\,\varphi_1(t)\,dt = \frac{1}{2\pi}\int_{-h}^{+h} \cos tx\,\varphi_1(t)\,dt$$

Puisque $\varphi_2(t)$ est une fonction caractéristique l'expression de gauche tend vers une valeur positive ou nulle quel que soit x. En particulier pour toutes les valeurs de x de la forme $\frac{2\pi}{2h}\,n$ on a :

$$A_n = \int_{-h}^{h} \cos \frac{2\pi}{2h} n t \;\varphi_1(t)\,dt \geqslant 0$$

Or les quantités $\frac{A_n}{h}$ sont les coefficients de Fourier de $\varphi_1(t)$ pris entre $-h$ et $+h$ pour les cosinus ; ceux relatifs aux sinus dans le même intervalle sont nuls à cause de la parité de $\varphi_1(t)$. Or $\varphi_1(t)$ est continue et monotone donc égale à la somme de sa série de Fourier.

Par conséquent :

$$\varphi_1(t) = \frac{A_0}{2h} + \sum_{n=1}^{+\infty} \frac{A_n}{h} \cos \frac{2\pi}{2h} t$$

ce qui montre que $\varphi_1(t)$ est fonction caractéristique. C'est la généralisation pour la loi de Polya de ce qu'avait établi Khintchine pour 1 - (t).

Le produit de deux fonctions caractéristiques de Polya est une fonction caractéristique de Polya. Mais l'axiome de simplification n'est pas valable.

Autrement dit si $\varphi_1, \varphi_2, \varphi_3$ sont trois fonctions caractéristiques de Polya l'égalité $\varphi_1\,\varphi_2 = \varphi_1\,\varphi_3$ n'entraine pas $\varphi_2 = \varphi_3$. Il suffit que φ_1 soit nul pour t supérieur en valeur absolue à h et que φ_2 et φ_3 coincident de -h à +h et différent ailleurs.

Donc l'ensemble des lois de Polya n'est pas un semi groupe au sens de l'algébre.

FORMATION DE f.c. NON CONVEXES A PARTIR DES FONCTIONS $k\left(\frac{t}{a}\right)$

Le produit de composition appliquée à des fonctions définies positives (f.d.p.) permet de construire de nouvelles f.d.p. (Voir M. GIRAULT note aux comptes rendus Ac. Sc. t 237, p 20-22, séance du 6 juillet 1953).

Un cas particulier de ces transformations permet d'énoncer le théorême suivant :

Théorême

Si $\varphi(t)$ est la f.c. réelle d'une distribution admettant partout une densité :

$$\Phi(t) = 2\,\varphi(t) + \varphi(t+h) + \varphi(t-h)$$

est f.d.p. quel que soit h, c'est-à-dire une f.c. à un coefficient positif près.

La densité de la répartition est donné par :

$$(4)\quad \int_{-\infty}^{+\infty} e^{-itx}\,\Phi(t)\,dt$$

or si $\displaystyle\int_{-\infty}^{+\infty} e^{-itx}\,\varphi(t)\,dt = r(x) \geqslant 0$

$$\int_{-\infty}^{+\infty} e^{-itx}\,\varphi(t+h)\,dt = e^{ihx}\int_{-\infty}^{+\infty} e^{-i(t+h)x}\,\varphi(t+h)\,d(t+h) = e^{ihx}\,r(x)$$

et de même $\displaystyle\int_{-\infty}^{+\infty} e^{-itx}\,\varphi(t-h)\,dt = e^{-ihx}\,r(x)$

L'intégrale (4) vaut donc $2\,r(x)\,(1+\cos hx) \geqslant 0$

Ces calculs ont, comme la démonstration de Polya une interprétation géométrique simple par des considérations de barycentres de masses $\varphi(t)$

Pour obtenir la densité qui correspond à $\Phi(t)$, il faut envisager le barycentre des masses $\varphi(t)$ enroulées sur le cercle trigonométrique.

Soit alors une valeur donnée de x ; $\varphi(t)$ a un barycentre M sur l'axe réel à une distance r = OM du centre. La répartition $\varphi(t+h)$ se déduit de la précédente par une rotation d'angle hx ; son barycentre M' est donc sur un même cercle que M centré à l'origine et $\varphi(t-h)$ a pour barycentre M" symétrique de M' toujours sur le même cercle. On est ramené à une répartition de 3 masses + 2 m en M, + m en M' et + m en M" dont le centre de gravité est sur la partie positive de l'axe réel.

LES CONTOURS POLYGONAUX DÉFINIS A PARTIR DES COEFFICIENTS DES BINOMES $(a+b)^{2n}$ SONT DES f.d.p.

On entend par là les fonctions f_n (t) définies par les valeurs suivantes

pour $|t| \geqslant n+1$ $\quad f_n = 0$

pour $t = \pm n$ $\quad f_n = C_{2n}^{0} = 1$

pour $t = \pm n-k$ $\quad f_n = C_{2n}^{k}$ $\quad$ k entier positif inférieur ou égal à n .

La fonction f_n est interpolée linéairement entre ces valeurs.

Soit la fonction F_0 de ce type définie par les valeurs 0,1,0 pour t = −1, 0,+ 1 c'est la fonction convexe k (t) et formons

$$F_1 = 2\ F_0(t) + F_0(t+1) + F_0(t-1)$$

elles prennent successivement les valeurs :

$F_0(t-1)$	0	1	0		
$2\,F_0(t)$		0	2	0	
$F_0(t+1)$			0	1	0
F_1	0	1	2	1	0

Puis à partir de F_1 on peut former :

$$F_2 = F_1(t-1) + 2\ F_1(t) + F_1(t+1)$$

dont les valeurs successives sont :

	0	1	2	1	0		
		0	2	4	2	0	
			0	1	2	1	0
F_2	0	1	4	6	4	1	0

coefficients du développement de $(a+b)^4$

On reconnait dans ce procédé la formation des coefficients successifs du binôme pour des ordres 2 n pairs, car en effet :

$$C_{n+2}^{p} = C_n^{p-2} + 2\ C_n^{p-1} + C_n^{p}$$

Donc tous ces contours polygonaux représentent des f.d.p. et si, par affinité on rend F (0) : 1 ce sont des fonctions caractéristiques, dont aucune à partir de F_2 n'est convexe.

Si n tend vers l'infini, la fonction $\varphi(t) = \frac{\sqrt{n\pi}}{2^{2n}}\, F_n \frac{t}{\sqrt{n}}$ tend vers la f.c. d'une loi normale. Ce qui démontre directement que e^{-t^2} est une fonction caractéristique.

COURBES ALÉATOIRES

par

Maurice FRÉCHET

Pour illustrer notre théorie générale des éléments aléatoires de nature quelconque (1) nous avions voulu faire une étude expérimentale de courbes aléatoires. Nous avions d'abord considéré à cet effet les formes que prenait sur un plan, un fil noué jeté au hasard sur ce plan. Nous avions l'intention d'en étudier une représentation paramètrique en séries de Fourier. Mais les formes obtenues pour cent jets du même fil se sont trouvées si compliquées, que le nombre de termes nécessaires dans le développement a dû dépasser 60. Il en résultait des calculs si longs que nous n'avons pu analyser ainsi les 100 courbes.

Nous avons donc momentanément abandonné cet essai et nous avons considéré des formes beaucoup plus simples, celles de contours craniens, pour lesquels une douzaine d'harmoniques assuraient une approximation suffisante. Mme LAURENT-DUHAMEL a bien voulu procéder à l'étude de ces contours et les résultats de son travail ont paru dans le vol.II., fasc.3 1953 du présent périodique.

Cependant, il nous a paru qu'il pourrait être utile de ne pas laisser ignorer tout ce qui avait été fait sur les fils aléatoires. Pour chaque jet on avait calqué la forme plane du fil jeté. Puis on a déterminé les courbes représentatives des coordonnées x, y d'un point de la courbe en fonction de son abscisse curviligne s. On a enfin déterminé pour plusieurs de ces courbes les 30 premières harmoniques des développement de x et y en série de Fourier en fonction de s.

Nous reproduisons ci-après les courbes représentatives de x et y pour une vingtaine de ces courbes.

---

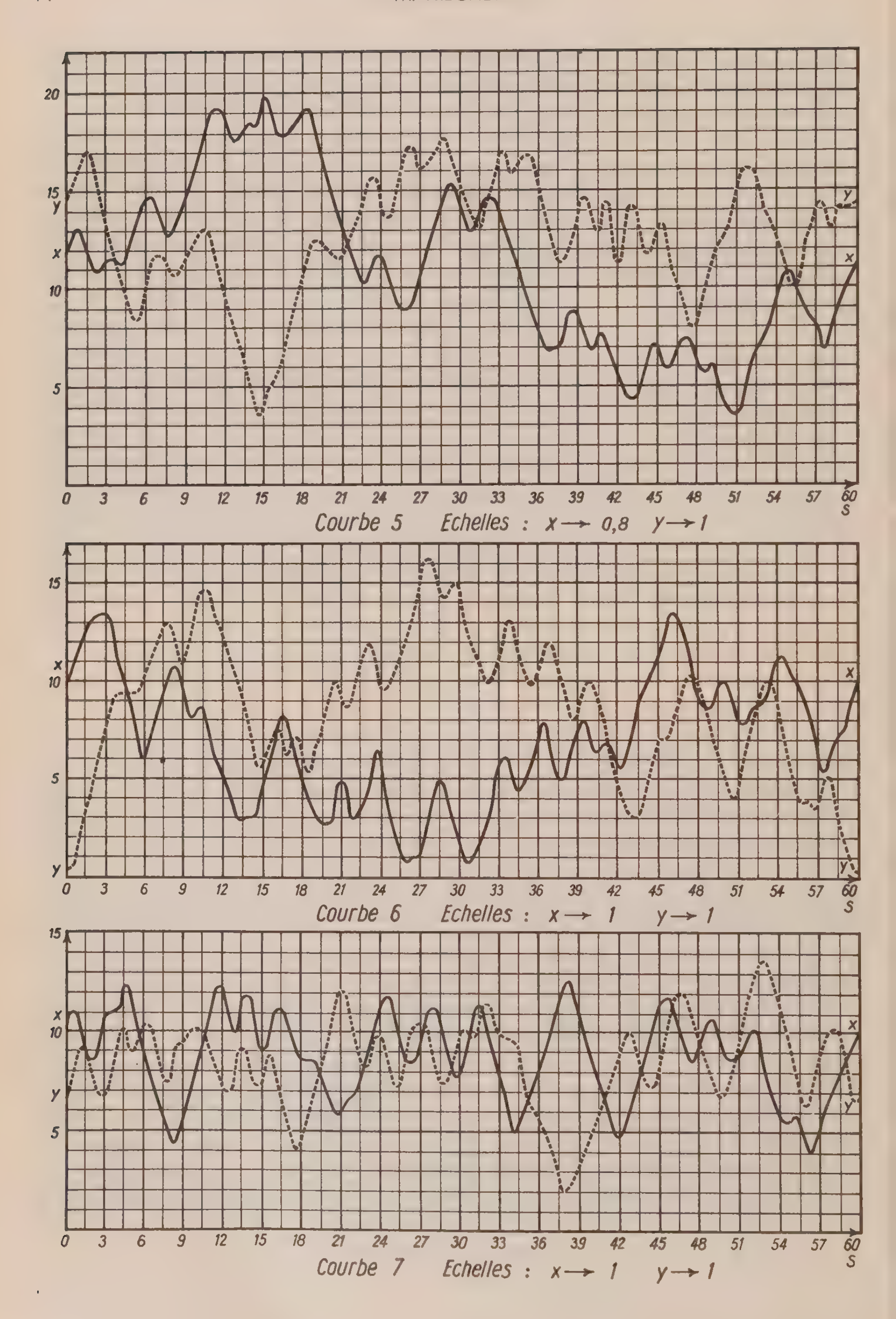
(1) Voir un résumé de cette théorie au Chapitre VI de la Deuxième édition du Premier Livre de nos Recherches théoriques modernes sur le Calcul des Probabilités, chez Gauthier Villars.

Courbe 1 Echelles $\begin{cases} x \longrightarrow 0{,}7 \\ y \longrightarrow 1 \end{cases}$

Courbe 2 Echelles $\begin{cases} x \longrightarrow 0{,}65 \\ y \longrightarrow 1 \end{cases}$

Courbe 3 Echelles { $x \longrightarrow \frac{2}{3}$; $y \longrightarrow 1$ }

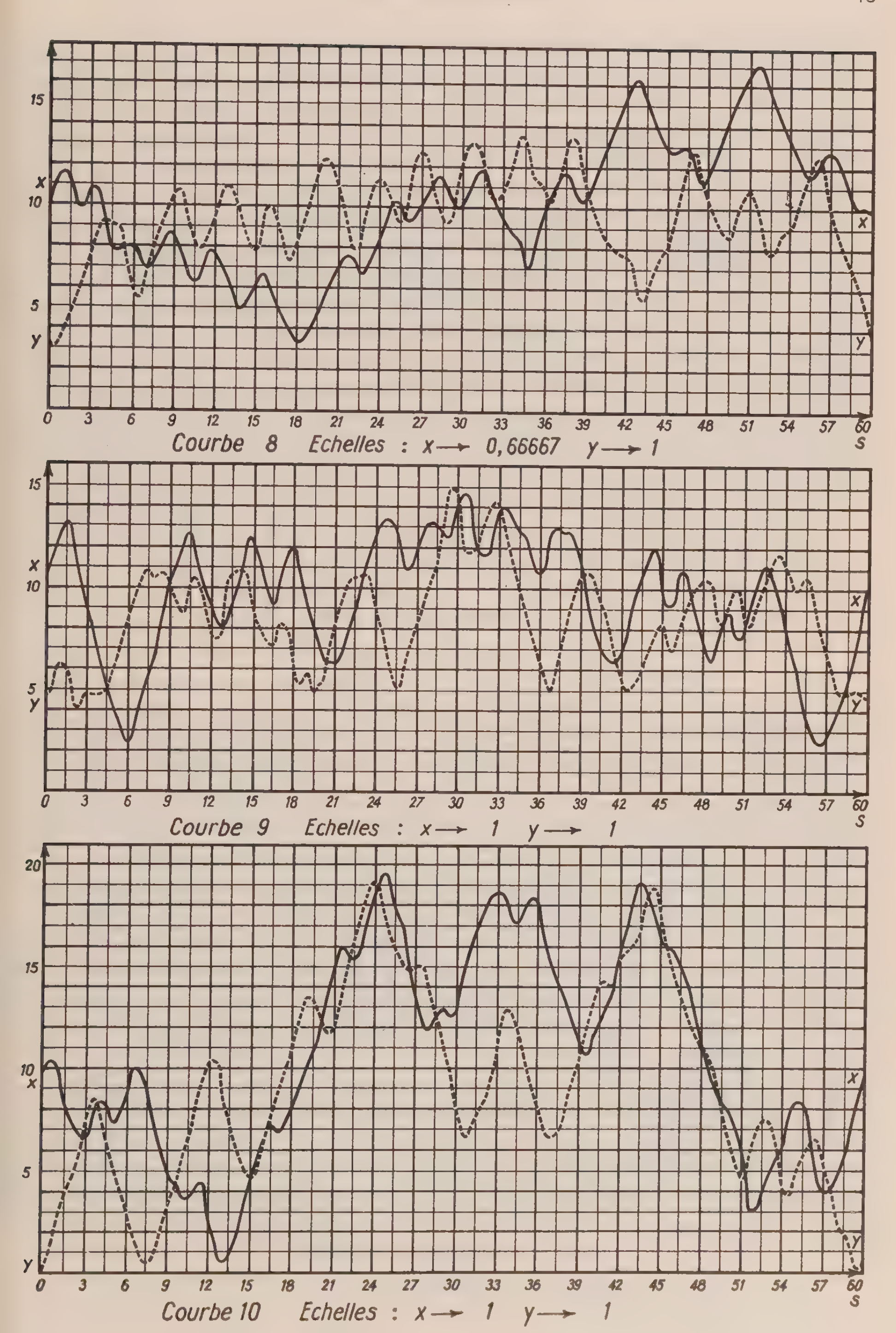
Courbe 4 Echelles { $x \longrightarrow 0{,}6$; $y \longrightarrow 1$ }



Courbe 5 Echelles : $x \to 0{,}8$ $y \to 1$

Courbe 6 Echelles : $x \to 1$ $y \to 1$

Courbe 7 Echelles : $x \to 1$ $y \to 1$



15
x
10
5
y
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
s
x
y
Courbe 8 Echelles : x ⟶ 0,66667 y ⟶ 1
15
x
10
5
y
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
s
x
y
Courbe 9 Echelles : x ⟶ 1 y ⟶ 1
20
15
10
x
5
y
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
s
x
y
Courbe 10 Echelles : x ⟶ 1 y ⟶ 1

Courbe 11 Echelles $\begin{cases} x \longrightarrow 1 \\ y \longrightarrow 1 \end{cases}$

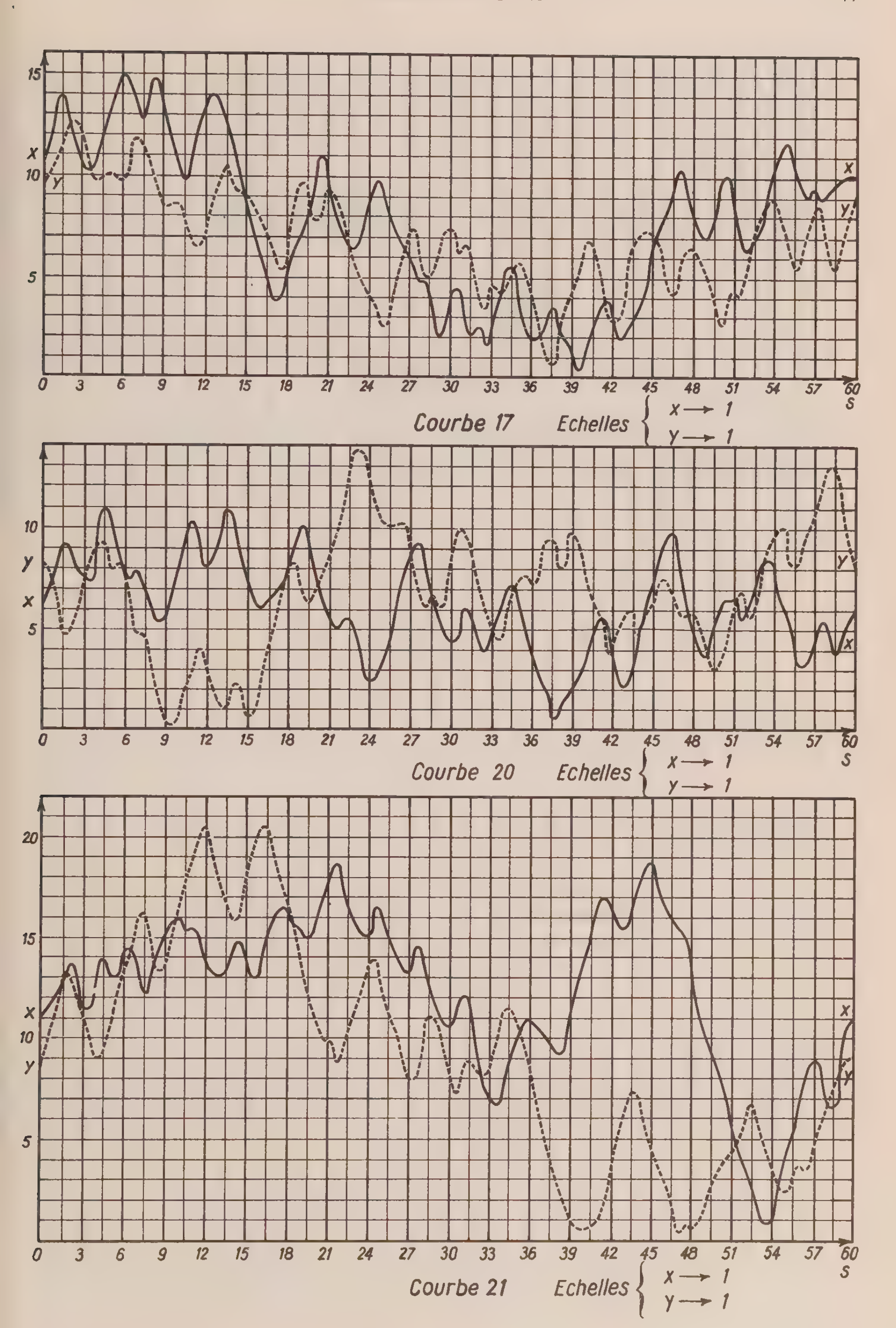
Courbe 12 Echelles $\begin{cases} x \longrightarrow 1 \\ y \longrightarrow 1 \end{cases}$

Courbe 13 Echelles $\begin{cases} x \to 1 \\ y \to 1 \end{cases}$

Courbe 14 Echelles $\begin{cases} x \to \frac{2}{3} \\ y \to 1 \end{cases}$

Courbe 15 Echelles $\begin{cases} x \to 1 \\ y \to 1 \end{cases}$

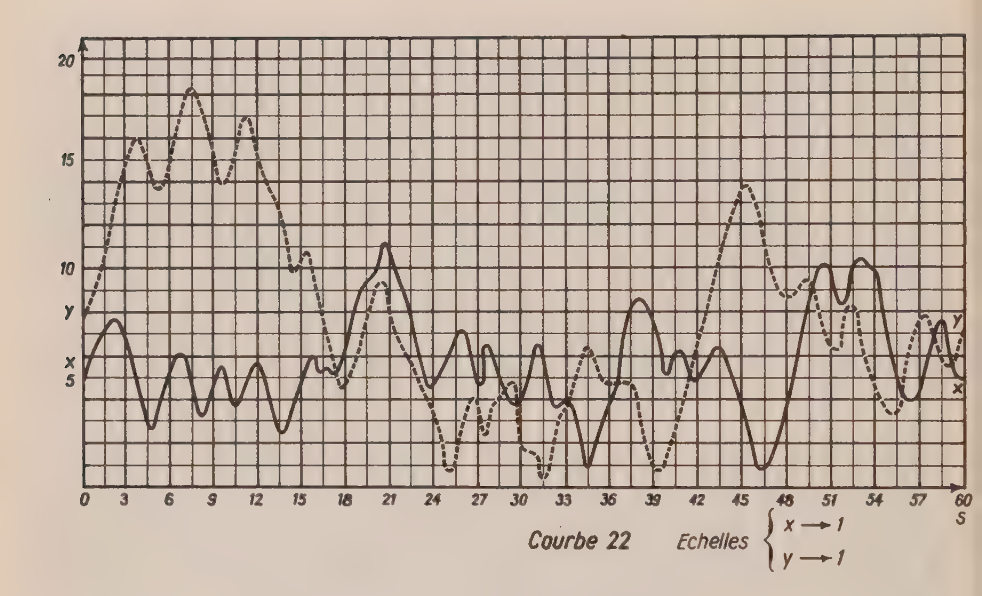
Courbe 16 Echelles $\begin{cases} x \to 1 \\ y \to 1 \end{cases}$



Courbe 17 Echelles { x → 1, y → 1 }

Courbe 20 Echelles { x → 1, y → 1 }

Courbe 21 Echelles { x → 1, y → 1 }


20
15
10
y
x
5
0
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
51
54
57
60
s
y
x
Echelles
x → 1
y → 1
Courbe 22

LES FONCTIONS FACTORIELLES

par

ETIENNE HALPHEN

AVANT-PROPOS

L'Académie des Sciences a décerné en 1954 à Etienne HALPHEN le prix Montyon de Statistique "pour ses travaux en hydrologie statistique". De notre grand ami disparu, les publications de l'I.S.U.P. accueillent aujourd'hui deux mémoires : l'un porte sur les fondements de la Théorie des Probabilités ("La notion de vraisemblance") - le second, malheureusement inachevé, contient l'étude d'une remarquable famille de fonctions, qu'Etienne HALPHEN avait rebaptisées "fonctions factorielles". Entre ces deux mémoires et l'hydrologie, le lien n'est ni direct ni très apparent. C'est pourtant à cette science attachante qu'Etienne HALPHEN était redevable des travaux publiés dans ce fascicule.

Ses conceptions sur la notion de "vraisemblance", comme il aimait à le faire observer, étaient liées étroitement à ceci; la chance d'avoir travaillé, en statistique, sur un matériel très pauvre et très imparfait (les observations hydrologiques et météorologiques). Et sans doute, le statisticien qui travaille dans un domaine où l'expérimentation est possible et bon marché, n'éprouvera pas la nécessité d'approfondir les bases logiques de ses méthodes : c'est qu'il peut se borner à utiliser une "information" soigneusement triée, de telle sorte que l'adéquation du calcul des probabilités ne pose même pas de question. Mais l'hydrologue qui voudrait se borner aussi à une telle information se retrouverait très pauvre - et même impuissant à résoudre aucun problème. Etienne HALPHEN n'accordait à des expressions telles que "méthode de cuisine", aucun sens péjoratif... au contraire : la bonne cuisine étant l'art de "corriger" la théorie probabiliste pour améliorer son adéquation aux choses concrètes.

Quant au mémoire qu'on va lire, il faut dire comment s'était imposée l'étude des fonctions "factorielles". Ayant constaté que dans l'arsenal des lois de probabilité couramment utilisées, depuis Pearson, par les statisticiens, l'hydrologue ne trouve pas toujours... chaussure à son pied, Etienne HALPHEN avait introduit de nouvelles familles, les types A et B, dont les densités s'expriment par les formules suivantes :

$$\text{type A : } f(x) = \frac{\mu^{\gamma}}{2K_{\gamma}(\alpha)} e^{-\frac{\alpha}{2}\left(\frac{x}{\mu} + \frac{\mu}{x}\right)} x^{\gamma-1} ,$$

$$\text{type B : } f(x) = \frac{1}{B(\alpha, \beta, \gamma)} e^{-\alpha\frac{x^2}{2} - \beta x} x^{\gamma-1} ,$$

Elles définissent deux familles de lois à trois paramètres, dont les formes sont suffisamment variées pour s'adapter aux besoins de l'hydrologie, de la météorologie, et peut-être aussi (il est permis de l'espérer) de quelques autres domaines. Les premiers résultats concernant l'application de ces lois à l'étude des débits ont été publiés dès 1949 ("Les lois de débits des rivières françaises", La Houille Blanche, numéro spécial B. 1949). Mais leur utilisation se trouvait très limitée parce que ces lois n'étaient pas tabulées (mises à part quelques formes particulières : lois de Pearson des types III et V, loi harmonique de Halphen, loi de χ^2 et loi de Gauss notamment).

La tabulation complète des lois A et B de Halphen exigeait l'emploi, pour les lois A des fonctions de Bessel-Hankel, pour les lois B des fonctions de Hermite; sur ces dernières, il semble que peu de travaux numériques aient été exécutés. Ce sont elles qu'Etienne HALPHEN étudie sous le nom nouveau de "fonctions factorielles".

Mais, comme on va le voir dans les pages suivantes, ce n'est pas le seul nom qui est nouveau. C'est aussi l'état d'esprit avec lequel Etienne HALPHEN aborde cette étude, et l'exposé fort élégant qu'il en donne. C'est encore plusieurs résultats assez remarquables.

Il ne parait pas qu'on ait vu avant lui une foule de liens qui font des "factorielles" un "carrefour important de la Théorie des fonctions ...".

Leur étude amène fort naturellement Etienne HALPHEN à l'identité :

$$\int_0^\infty e^{-z^2}(\cos xz + \sin xz)\frac{dz}{\sqrt{z}} \equiv \int_0^\infty e^{-(z-\frac{x}{2})^2}\frac{dz}{\sqrt{z}}$$

qui semble nouvelle.

Et ajoutons, pour clore cet avant-propos, que la tabulation complète des lois A et B est aujourdh'ui en cours d'achèvement.

G. MORLAT

LES FONCTIONS FACTORIELLES

par

ETIENNE HALPHEN

GÉNÉRALITÉS

A - DÉFINITIONS

Nous donnerons les noms de "cosinus factoriel" et "sinus factoriel" aux deux fonctions définies par les séries suivantes(1) :

$$(1)\quad cf_\alpha (x) = 1 + \alpha \frac{x^2}{2!} + \alpha(\alpha + 1) \frac{x^4}{4!} + \ldots\ldots$$

$$(2)\quad sf_\alpha (x) = x + \left(\alpha + \frac{1}{2}\right) \frac{x^3}{3!} + \left(\alpha + \frac{1}{2}\right)\left(\alpha + \frac{3}{2}\right) \frac{x^5}{5!} + \ldots$$

Ces séries convergent pour toutes valeurs (réelles ou imaginaires) de x et de α, en sorte que les fonctions cf et sf sont fonctions entières de x - et aussi fonctions entières de α, comme on le voit en dérivant terme à terme en α.

En posant : $x = iz$, on est conduit aux fonctions suivantes (dans lesquelles nous préférons désigner par une autre lettre, λ par exemple, le "paramètre") :

$$(3)\quad cg_\lambda (z) = 1 - \lambda \frac{z^2}{2!} + \lambda(\lambda + 1) \frac{z^4}{4!} \ldots\ldots$$

$$(4)\quad sg_\lambda (z) = z - \left(\lambda + \frac{1}{2}\right) \frac{z^3}{3!} + \left(\lambda + \frac{1}{2}\right)\left(\lambda + \frac{3}{2}\right) \frac{z^5}{5!} \ldots$$

Nous donnons aux fonctions : cf, sf, et aux fonctions apparentées, le nom de "fonctions factorielles de première espèce"; et celui de "fonctions factorielles de seconde espèce" à celles qui s'apparentent à cg et sg.

On voit immédiatement que :

$$(5)\quad cf'_\alpha = \alpha\, sf_{\alpha+\frac{1}{2}} \qquad\qquad (7)\quad cg'_\lambda = -\lambda\, sg_{\lambda+\frac{1}{2}}$$

$$(6)\quad sf'_\alpha = cf_{\alpha+\frac{1}{2}} \qquad\qquad (8)\quad sg'_\lambda = cg_{\lambda+\frac{1}{2}}$$

en notant, de façon abrégée, par cf' la dérivée de cf par rapport à la "variable" x, etc .... (2)

---

(1) Lire : " cf de x ", etc...

(2) De façon générale, nous faisons usage de pareilles abréviations, où nous sous-entendons la "variable" ou le "paramètre", chaque fois qu'il n'y a pas risque d'équivoque.

La combinaison linéaire :

(9) $ef_{\alpha}(x) = cf_{\alpha} \cdot \Gamma(\alpha) + sf_{\alpha} \cdot \Gamma(\alpha + \frac{1}{2})$

ou

(10) $ef_{\alpha}(x) = \Gamma(\alpha) + \Gamma(\alpha+\frac{1}{2}) \cdot \frac{x}{1!} + \Gamma(\alpha+1) \cdot \frac{x^2}{2!} \ldots$

donne lieu à la formule de dérivation :

(11) $ef'_{\alpha} = ef_{\alpha+\frac{1}{2}}$,

ce qui nous conduit à la nommer : "exponentielle factorielle" de première espèce. Nous nommerons "exponentielle factorielle" de seconde espèce la combinaison linéaire :

(12) $eg_{\lambda} = cg_{\lambda} \cdot E(\lambda) + sg_{\lambda} \cdot E(\lambda + \frac{1}{2})$,

où $E(\lambda)$ désigne la fonction entière : $\frac{1}{\Gamma(1-\lambda)}$.

La raison de cette dénomination se trouve dans la formule de dérivation :

$$(13) \quad eg'_{\lambda} = eg_{\lambda+\frac{1}{2}} .$$

Notons encore :

(14) $eg_{\lambda}(x) = E(\lambda) + \frac{x}{1!} E(\lambda + \frac{1}{2}) + \ldots.$

ainsi que la formule :

(15) $cf_{\alpha}(x) = \frac{1}{2\Gamma(\alpha)} \left[ef_{\alpha}(x) + ef_{\alpha}(-x) \right]$

qui, avec les formules similaires, ramène l'étude de cf et sf à celle de ef, etc...

Il est important de noter que, à la différence de ef, la fonction eg est fonction entière du paramètre :
nous avons des raisons de penser que ce détail, en apparence insignifiant, est ce qui donne une importance exceptionnelle.

L'équation fonctionnelle :

$$(F) \quad \frac{\partial}{\partial x} f(x, r) = f(x, r + a)$$

semble en effet une "clé" de l'étude des fonctions analytiques, le développement en série entière d'une fonction analytique étant liè à cette particularité de la fonction monôme $\frac{x^r}{r!}$ de vérifier l'équation (F). Mais ce n'est pas une fonction entière en r : d'ou les singularités que l'on rencontre en particulier dans l'intégration de $\frac{x^r}{r!}$ pour $r = -1$; à noter, plus généralement, que la fonction monôme n'est même pas une fonction uniforme en x pour r non entier, tandis que $ef_{\alpha}(x)$ est uniforme pour toute valeur de α pour laquelle elle existe, et que $eg_{\lambda}(x)$ est même, comme nous venons de le faire remarquer, fonction entière de x et de λ .

B - PROPRIÉTÉS ÉLÉMENTAIRES

De la formule (10), on tire la combinaison :

$$\mathrm{ef}_{\alpha+1}(x) - \frac{x}{2}\cdot \mathrm{ef}_{\alpha+\frac{1}{2}} = \Gamma(\alpha+1) + \frac{x}{1!}\left[\Gamma(\alpha+\tfrac{3}{2}) - \tfrac{1}{2}\,\Gamma(\alpha+\tfrac{1}{2})\right] + \ldots$$
$$= \alpha\left[\Gamma(\alpha) + \frac{x}{1!}\,\Gamma(\alpha+\tfrac{1}{2}) + \ldots\right]$$

ou

(16) $$\mathrm{ef}_{\alpha+1} = \frac{x}{2}\,\mathrm{ef}_{\alpha+\frac{1}{2}} + \alpha\,\mathrm{ef}_{\alpha}\,.$$

En vertu de (11), cette équation (16) montre que la fonction de x : $y = \mathrm{ef}(x)$ est solution de l'équation différentielle :

(E) $$y'' = \frac{x}{2}\,y' + \alpha y\,.$$

On reconnait à un changement de variable près, l'équation classique vérifiée, lorsque 2α est entier négatif, par le polynôme de Hermite d'ordre -2α.

On trouvera de même que $u = \mathrm{eg}_{\lambda}(z)$ est solution de l'équation différentielle :

(E') $$u'' + \frac{z}{2}\,u' + \lambda u = 0\,.$$

La solution générale de (E) peut donc normalement s'écrire :

(17-18) $$y = A\,\mathrm{cf}_{\alpha} + B\,\mathrm{sf}_{\alpha} = C\,\mathrm{ef}_{\alpha}(x) + D\,\mathrm{ef}_{\alpha}(-x)$$

et celle de (E') :

(19-20) $$u = A\,\mathrm{cg}_{\lambda} + B\,\mathrm{sg}_{\lambda} = C\,\mathrm{eg}_{\lambda}(z) + D\,\mathrm{eg}_{\lambda}(-z)\,.$$

Nous aurions pu <u>définir</u> les fonctions factorielles comme étant les solutions des équations (E) et (E') (1) ; en ce cas, cf serait par définition la solution vérifiant pour $x = 0$ les conditions initiales de Cauchy : $y = 1$, $y' = 0$, etc. Cette remarque donne immédiatement la démonstration des relations remarquables suivantes entre fonctions factorielles de première et de seconde espèce :

(21) $$\mathrm{cg}_{\lambda}(x) = e^{-\frac{x^2}{4}}\cdot \mathrm{cf}_{-\lambda+\frac{1}{2}}(x)\,,$$

(22) $$\mathrm{sg}_{\lambda}(x) = e^{-\frac{x^2}{4}}\cdot \mathrm{sf}_{-\lambda+\frac{1}{2}}(x)\,.$$

Pour démontrer ces formules, il suffit de vérifier que les seconds membres satisfont, en raison de (E), à l'équation (E'); les conditions initiales de Cauchy pour $x = 0$ achèvent la démonstration.

On notera encore que :

(23) $$\mathrm{eg}_{\lambda}(x) = \frac{1}{\Gamma(\frac{1}{2}-\lambda)\,\Gamma(1-\lambda)}\,e^{-\frac{x^2}{4}}\cdot \mathrm{ef}_{-\lambda+\frac{1}{2}}(x)$$

(même démonstration).

(1) Nous avons préféré suivre l'ordre précédent pour faire l'économie de discussions fastidieuses.

C - LES FONCTIONS FACTORIELLES HOMOGÈNES

Les équations différentielles (E), (E'), linéaires et homogènes du second ordre, peuvent être ramenées à des équations du premier ordre en prenant pour nouvelles fonctions les dérivées logarithmiques des y ou des u. Cela conduit à étudier spécialement ces dérivées logarithmiques, qui présentent, en raison des équations (5), (6), (7) et (8), un intérêt assez notable. Exposons les calculs dans le cas des fonctions de première espèce :

Il est commode de poser :

(24) $$af_\alpha(x) = \frac{cf_\alpha}{sf_{\alpha+\frac{1}{2}}}$$

(25) $$bf = \frac{cf_{\alpha+\frac{1}{2}}}{sf_\alpha}$$

D'après (E), compte tenu de (5) et (6), on obtient immédiatement :

(26) $$bf_{\alpha+\frac{1}{2}}(x) - af_\alpha(x) = \frac{x}{2}.$$

Mais, d'autre part, en prenant la dérivée logarithmique des deux membres de (24), on obtient :

$$\frac{af'_\alpha}{af_\alpha} = \frac{\alpha\, sf_{\alpha+\frac{1}{2}}}{cf_\alpha} - \frac{cf_{\alpha+1}}{sf_{\alpha+\frac{1}{2}}}$$

Ou :

$$af'_\alpha + af_\alpha \,.\, bf_{\alpha+\frac{1}{2}} = \alpha$$

qui donne, en tenant compte de (26) :

$$af'_\alpha + af_\alpha\left(af_\alpha + \frac{x}{2}\right) = \alpha.$$

Ainsi la fonction $af_\alpha(x)$ est solution de l'équation de Riccati :

(R) $$z' + z\left(z + \frac{x}{2}\right) = \alpha.$$

En posant : $bf_\alpha(x) = t$, on obtient, en tenant compte de (26) dans l'équation (R) :

(S) $$t' + t\left(t - \frac{x}{2}\right) = \alpha$$

En outre, les fonctions :

$$uf_\alpha = \frac{\alpha}{af_\alpha} \quad \text{et} \quad vf_\alpha = \frac{\alpha}{bf_\alpha}$$

vérifient encore, respectivement, les équations :

(R) pour : $z = vf_\alpha(x)$; et (R') pour : $t = uf_\alpha(x)$.

On trouve sans difficulté que l'intégrale générale de (R) est : $\frac{\alpha\, y}{y'}$, y étant l'intégrale générale de (E) tandis que $\frac{y'}{y}$ est l'intégrale générale de l'équation (S).

Si l'on part de l'équation de seconde espèce (E'), on obtient de la même façon les solutions des équations de Riccati :

$$(R') \quad z' - z\left(z + \frac{x}{2}\right) = \lambda$$

et

$$(S') \quad t' - t\left(t - \frac{x}{2}\right) = \lambda .$$

Nous donnerons aux fonctions : af_{α} ... ; et : ag_{λ} .... : les noms de "fonctions factorielles homogènes" de première et de seconde espèce respectivement.

INTÉGRALES DÉFINIES

- A -

Nous groupons dans ce chapitre des expressions des fonctions factorielles par des intégrales ; et d'abord

$$(33) \qquad ef_\alpha (x) = 2 \int_0^\infty e^{-z^2 + xz} \cdot z^{2\alpha-1} \cdot dz , \qquad (\alpha > 0) .$$

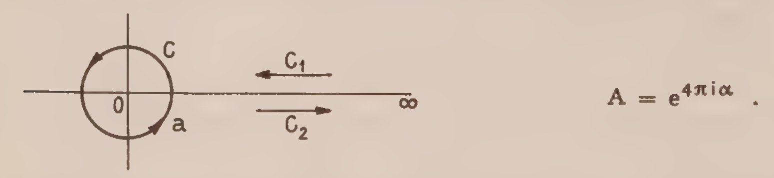
La formule (33) s'établit en développant e^{xz} en série entière, et intégrant terme à terme le produit par :

$$e^{-z^2} \cdot z^{2\alpha-1} \cdot dz .$$

La démonstration rigoureuse peut se faire de plusieurs façons ; par exemple en intégrant la série dans l'intervalle (o, a), et prenant une majorante de l'intégrale $\int_a^\infty$.

Nous aurons à détailler un autre mode de démonstration, pour obtenir un développement asymptotique de ef .

La formule (33) n'a de sens que pour $\alpha > 0$. Si α est négatif, ou s'il est imaginaire, on peut obtenir pour ef une intégrale de Cauchy de la façon suivante : $z^{2\alpha-1}$ envisagé comme fonction de la variable complexe Z a un point de ramification à l'origine : les déterminations qui s'échangent par un parcours simple fermé autour de O étant différentes, pourvu que $2\alpha - 1$ ne soit pas entier. Dès lors, pour 2α non entier, mais >0, les intégrales de Cauchy : $-\int_\infty^a$ et $\int_a^\infty$ prises sur l'axe réel sont différentes si, avant de passer de la première à la seconde, on décrit un circuit autour de O; la seconde se déduit de la première en la multipliant par le facteur :



$$A = e^{4\pi i\alpha} .$$

Pour α réel > 0, lorsque $a \longrightarrow 0$, l'intégrale autour de l'origine tend vers zéro, puisqu'on peut toujours prendre pour C un cercle de rayon a . Donc, l'intégrale prise sur le parcours : (C_1, C, C_2) et qui ne dépend pas de a, est égale à l'intégrale (33), au facteur $(A-1)$ près.

Comme au reste c'est une fonction monogène et uniforme de α pour toute valeur de α, elle définit dans tout le plan des α une fonction analytique uniforme, qui réalise donc le prolongement analytique de (33), et donne ef_α (x) lorsque $A - 1 \neq 0$; pour $A - 1 = 0$, on retrouverait les pôles en α de ef_α (il suffit, pour le démontrer, de faire tendre α vers un tel pôle).

Ainsi nous pouvons, quel que soit α , écrire

(34) $$\int_{(C_1, C, C_2)} e^{-z^2+xz} \, . \, z^{2\alpha-1} \, . \, dz = \frac{e^{4\pi i\alpha}-1}{2} \, ef_\alpha (x) \, .$$

On observera que pour $x = O$, on retrouve l'intégrale de Cauchy pour la fonction Γ .

\- B -

Ainsi l'unique fonction de deux variables $ef(x, \alpha)$ qui suffit, moyennant quelques précautions, a donner toutes les fonctions factorielles par des opérations élémentaires, fournit comme cas particuliers ou cas limites :

- la fonction exponentielle;
- les fonctions trigonomètriques;
- la fonction eulérienne;
- les polynômes (et les fonctions) de Hermite;

tout cela (sauf les fonctions de Hermite) sans sortir, pour x et α , du domaine du réel.

On achèvera dans la suite, de se rendre compte que toutes les formules classiques relatives aux fonctions ci-dessus, se rattachent à des formules ou propriétés générales de la fonction ef .

On peut en dire autant de l'unique fonction de deux variables :

$$\boxed{eg(x, \lambda)}$$

étant entendu que les précautions à prendre (1) pour $\lambda = -\frac{n}{2}$ (n entier $> O$) sont un peu différentes de celles qui concernent $ef(x, \alpha)$ pour $\alpha = -\frac{n'}{2}$ (n' entier $\geqslant O$) : et nous avons déja dit l'intérêt exceptionnel de la fonction eg qui est entière en x et en λ , et admet, quel que soit λ , une primitive en x et une seule qui soit une fonction eg .

C'est ici le lieu d'expliquer l'intérêt du domaine réel dans l'étude des fonctions factorielles :

L'extension au domaine complexe est un procédé (= prolongement analytique) pour substituer à l'étude d'une fonction d'une variable (mettons e^x) une fonction de deux variables (savoir : $e^x \operatorname{Cos} y$ - puisqu'aussi bien, comme Emile Picard le soulignait, l'étude d'une fonction analytique est équivalente à celle d'une fonction harmonique (2). Et l'étude d'une fonction de deux variables est plus "harmonieuse" que celle d'une fonction d'une variable, parce que c'est l'étude d'une classe à un paramètre continu de fonctions d'une variable : nous sommes ici en présence d'un fait bien connu en théorie des groupes (3) et que nous pensons rendre plus clair ultérieurement par des remarques sur le calcul opérationnel. Dès maintenant relevons que l'intérêt de la classe ici étudiée se rattache au fait qu'elle se concerve par l'opérateur $\frac{d}{dx}$ qui équivaut ici à un opérateur linéaire appliqué au paramètre.

(1) En vue d'obtenir toutes les fonctions factorielles, solutions de E'.

(2) Il serait bon, dans la définition des fonctions monogènes, de parler de fonction réelle ou imaginaire : $P(z) + i\,Q(z)$ d'une variable complexe - les conditions de Cauchy pouvant s'écrire $\frac{df}{dx} = \frac{df}{diy}$ où il importe peu que f soit réelle ou imaginaire.

(3) Et en calcul fonctionnel.

En prenant les parties paire et impaire de (34), on obtient :

(35) $$\int_{C_1 \subset C_2} e^{-z^2} \operatorname{ch}(xz) \cdot Z^{2\alpha-1} dz = \frac{e^{4\pi i\alpha}-1}{2} \Gamma(\alpha)\, cf_\alpha(x)$$

(36) $$\int_{C_1 \subset C_2} e^{-z^2} \operatorname{sh}(xz) \cdot Z^{2\alpha-1} dz = \frac{e^{4\pi i\alpha}-1}{2} \Gamma\left(\alpha+\tfrac{1}{2}\right) sf_\alpha(x)$$

D'où, en changeant x en ix :

(37) $$\int_{C_1 \subset C_2} e^{-z^2} \cos(xz) \cdot Z^{2\lambda-1} dz = \frac{e^{4\pi i\lambda}-1}{2} \Gamma(\lambda)\, cg_\lambda(x) .$$

(38) $$\int_{C_1 \subset C_2} e^{-z^2} \sin(xz) \cdot Z^{2\lambda-1} dz = \frac{e^{4\pi i\lambda}-1}{2} \Gamma\left(\lambda+\tfrac{1}{2}\right) sg_\lambda(x)$$

En multipliant (37) par $\dfrac{1}{\Gamma(\lambda)\Gamma(1-\lambda)} = \dfrac{\sin\pi\lambda}{\pi}$

et (38) par $\dfrac{1}{\Gamma\left(\frac{1}{2}+\lambda\right)\Gamma\left(\frac{1}{2}-\lambda\right)} = \dfrac{\cos\pi\lambda}{\pi}$

on obtient

(39) $$\int_{C_1 \subset C_2} e^{-z^2} \left[\cos xz \cdot \sin\pi\lambda + \sin xz \cdot \cos\pi\lambda\right] z^{2\lambda-1}\, dz = \pi\, \frac{e^{4\pi i\lambda}-1}{2} eg_\lambda(x)$$

Les intégrales de Cauchy (35) à (39) donnent des intégrales de Riemann lorsque α ou λ sont réels > 0. En particulier :

(40) $$eg_\lambda(x) = \frac{2}{\pi} \int_0^\infty e^{-z^2} \sin(xz+\lambda\pi) \cdot z^{2\lambda-1}\, dz .$$

Lorsque $0 < \lambda < \frac{1}{2}$, la relation (23) entre eg et ef permet leur expression par deux intégrales de Riemann différentes. Il en résulte une identité remarquable entre ces deux intégrales; la voici dans le cas particulièrement frappant où $\lambda = \alpha = \frac{1}{4}$ (1)

(41) $$\int_0^\infty e^{-z^2} (\cos xz + \sin xz) \frac{dz}{\sqrt{z}} = \int_0^\infty e^{-\left(z-\frac{x}{2}\right)^2} \frac{dz}{\sqrt{z}} .$$

La formule de "réflexion" de Γ au point $\frac{1}{2}$ apparait donc elle aussi comme un cas particulier d'une propriété des fonctions factorielles (mais les unités que nous avons adoptées (2) donnent à ce point l'affixe $\frac{1}{4}$).

---

(1) Une démonstration directe de cette identité particulière a été donnée dans les Comptes Rendus de l'Académie des Sciences, séance du 23 novembre 1953.

(2) Ce n'est pas pour le plaisir de compliquer le raccordement avec les notations traditionnelles que nous avons adopté ces unités : c'est pour la simplicité des séries qui définissent les fonctions factorielles : (1) etc...

DÉVELOPPEMENTS ASYMPTOTIQUES

Nous limiterons notre exposé aux cas où l'on peut exprimer les fonctions factorielles par intégrales de Riemann : cela nous suffira en pratique. La considération des intégrales de Cauchy présente des complications, mais ne donne pas lieu à grandes difficultés théoriques.

Nous supposons donc α ou λ réels > 0; nous étudierons ef_α et eg_λ pour x réel, et chercherons leurs comportements :

pour : $x \longrightarrow +\infty$, et : $x \longrightarrow -\infty$;

ainsi que pour : $\alpha \longrightarrow +\infty$, et : $\lambda \longrightarrow +\infty$

- A - $\boxed{x \longrightarrow \pm\infty}$

1°) - Pour $x \longrightarrow +\infty$, avec $\alpha > 0$, nous procéderons de la façon suivante : on a, d'après (33) :

$$ef_\alpha(x) = 2\int_0^\infty e^{-z^2+zx}\, z^{2\alpha-1}\,dz = 2\,e^{\frac{x^2}{4}}\int_{-\frac{x}{2}}^\infty e^{-u^2}\left(u+\frac{x}{2}\right)^{2\alpha-1} du \sim$$

$$\sim 2e^{\frac{x^2}{4}}\int_{-\infty}^{+\infty} e^{-u^2}\left(\frac{x}{2}\right)^{2\alpha-1}\left[1+\frac{(2\alpha-1)}{1!}\,\frac{2u}{x}+\frac{(2\alpha-1)\,(2\alpha-2)}{2!}\left(\frac{2u}{x}\right)^2+\ldots\right]du$$

en vertu du comportement asymptotique de la courbe de Gauss.

Le second membre est une somme d'intégrales calculables : ce n'est pas une série convergente, mais la somme des p premiers termes pairs fournit un développement limité en au voisinage de $x \longrightarrow +\infty$ (se voit en prenant une majorante du $(p+1)$ème terme.

Donc :

$$(42)\quad ef_\alpha(x) \sim 2\sqrt{\pi}\; e^{\frac{x^2}{4}}\;\left(\frac{x}{2}\right)^{2\alpha-1}\left[1+(2\alpha-1)\,(2\alpha-2)\,\frac{1}{x^2}\ldots\right]$$

Les termes suivants se calculent aisément; nous ne les explicitons pas, car nous n'en aurons guère besoin.

2°) - Pour $x \longrightarrow -\infty$, ($\alpha$ étant toujours > 0) on obtient un développement asymptotique de ef en développant dans l'intégrale (33) le terme e^{-z^2} :

$$ef_\alpha(x) \sim 2\int_0^\infty e^{xz}\, z^{2\alpha-1}\left(1-\frac{z^2}{1!}+\frac{z^4}{4!}\ldots\right)dz$$

(on le prouvera en prenant une majorante du reste).

Le second membre est une somme d'intégrales eulériennes, ce qui donne finalement :

$$\text{(43)} \qquad ef_{\alpha}(x) \sim 2\Gamma(2\alpha) . \frac{1}{|x|^{2\alpha}} \left[1 - \frac{2\alpha(2\alpha+1)}{1!} \frac{1}{x^2} + \dots \right]$$

3°) - α (ou λ) quelconque.

Nous avons obtenu des développements asymptotiques de ef_{α} au voisinage de $x = +\infty$ et de $x = -\infty$ lorsque α est positif. Un tel développement vérifie nécessairement une relation de récurrence entre trois valeurs de α décalées de $1/2$, en raison de la relation de récurrence (16) entre fonctions ef. Partant alors de deux valeurs positives de l'indice : $\alpha + 1$ et $\alpha + \frac{1}{2}$, on obtient les développements asymptotiques correspondant à α, et en remontant ainsi de $-\frac{1}{2}$ en $-\frac{1}{2}$, on obtiendra les développements correspondant à n'importe quelle valeur négative de α.

Les formes analytiques des développements (42) et (43) donnent lieu à des identités dans la relation de récurrence (1) : donc elles subsistent sans modification pour $\alpha < 0$, pourvu que α ne soit pas multiple entier de $-\frac{1}{2}$, (auquel cas d'ailleurs ef n'existe pas).

Lorsque α est multiple entier de $-\frac{1}{2}$, $\lambda = \frac{1}{2} - \alpha$ est multiple entier de $\frac{1}{2}$, et eg_{λ} se réduit à un polynôme multiplié par $e^{-\frac{x^2}{4}}$.

Récrivons les développements asymptotiques (42) et (43) sur la fonction eg :

$$\text{(44)} \qquad eg_{\lambda} \sim \frac{2\sqrt{\pi}}{\Gamma(\frac{1}{2}-\lambda)\,\Gamma(1-\lambda)} \left(\frac{2}{x}\right)^{2\lambda} \left[1 + 2\lambda(2\lambda+1)\frac{1}{x^2} + \dots \right]$$

pour $x \longrightarrow +\infty$

$$\text{(45)} \qquad eg_{\lambda} \sim \frac{2\Gamma(1-2\lambda)}{\Gamma\left(\frac{1}{2}-\lambda\right)\Gamma(1-\lambda)} e^{-\frac{x^2}{4}} (-x)^{2\lambda-1} \left[1 - \frac{(2\lambda-2)(2\lambda-1)}{1!} \frac{1}{x^2} + \dots \right]$$

pour $x \longrightarrow -\infty$

Formules non valables pour $\lambda = \frac{k}{2}$ (k entier positif).

- B - $\boxed{\alpha \longrightarrow +\infty}$

L'idée de la méthode, valable pour d'autres types d'intégrales, réside dans le fait que, si $f(z)$ est une fonction unimodale, $f(z)^{\alpha}$ tend à prendre une forme gaussienne (à des dilatations près) lorsque $\alpha \longrightarrow \infty$. Notre méthode permet par exemple une démonstration très simple de la formule de Stirling.

Pour alléger l'écriture, désignons par $\omega(x, \alpha)$ le mode de la fonction qui figure sous le signe $\int$ dans (33) :

$$\omega = \frac{x}{4} + \sqrt{\alpha - \frac{1}{2} + \frac{x^2}{16}}$$

---

(1) Et la relation (16) permet de vérifier que le reste d'ordre n tend vers 0 avec $\frac{1}{|x|}$.

Posons alors :

$$z = \omega(x, \alpha) + v .$$

En faisant "sortir" le plus possible de l'intégrale (33) on peut écrire :

$$ef_{\alpha}(x) = 2\omega^{2\alpha-1}.e^{\omega x-\omega^2} \int_{-\omega}^{+\infty} e^{-v^2+(x-2\omega)v} \left(1+\frac{v}{\omega}\right)^{2\alpha-1} dv .$$

En prenant $1+\frac{v}{\omega}$ comme l'exponentielle de son logarithme, et développant ce dernier, on a :

$$\left(1+\frac{v}{\omega}\right)^{2\alpha-1} = e^{(2\alpha-1)\left(\frac{v}{\omega}-\frac{v^2}{2\omega^2}\right)+\varepsilon}$$

ou ε tend vers O lorsque α et $\omega \longrightarrow \infty$. La partie principale de l'intégrale est une intégrale de Gauss dont l'écart-type est infiniment petit lorsque α et $\omega \longrightarrow \infty$; on peut donc confondre la borne inférieure d'intégration avec $-\infty$, et l'on trouve finalement

$$ef_{\alpha}(x) = \frac{2\sqrt{\pi}\,.\,\omega^{2\alpha}}{\sqrt{\omega+\alpha-\frac{1}{2}}} \cdot e^{-\omega^2+\omega x}\left[1+O(\omega)\right]$$

où $O(\omega)$ est de l'ordre de $\frac{1}{\omega}$.

En tenant compte de $\alpha-\frac{1}{2}=\omega^2-\frac{\omega x}{2}$, on obtient l'harmonieuse formule :

(46) $$ef_{\alpha}(x) \sim \frac{\left(\frac{\omega^2}{e}\right)^{\omega^2}}{\left(\frac{\omega}{e}\right)^{\omega x}} \cdot \sqrt{2\pi}$$

Pour $x = O$, on obtient un équivalent de la formule de Stirling $(\alpha = \omega^2+\frac{1}{2})$.

Dans la pratique, nous aurons besoin d'expliciter la formule en α . On obtient, en posant $\beta^2 = \alpha - \frac{1}{2}$:

(47) $$ef_{\alpha}(x) \sim \left(\frac{\beta^2}{e}\right)^{\beta^2} e^{\beta x+\frac{x^2}{8}} \sqrt{2\pi}$$

Il sera commode d'utiliser ce résultat sous la forme suivante, grâce à la formule (équivalente à celle de Stirling) qui s'obtient en faisant $x = O$:

(48) $$ef_{\alpha}(x) \sim \Gamma(\alpha)\ e^{\frac{x^2}{8}+x\sqrt{\alpha-\frac{1}{2}}}$$

Cette formule, exacte pour $x = O$, est d'autant plus précise que x est petit devant α .

- C - $\boxed{\lambda \longrightarrow +\infty}$

Le calcul est analogue, mais plus simple. On pose encore $z = \omega + v$, mais on laisse ω provisoirement indéterminé, et par la suite on est conduit à prendre, dans l'intégrale (40) :

$$\omega^2 = \lambda - \frac{1}{2} \cdot$$

Le calcul donne séparément cg_λ et sg_λ ; pour eg_λ on trouve

(49) $$eg_\lambda (x) \sim R(\lambda) \cdot \left(\frac{\omega^2}{e}\right)^{\omega^2}$$

avec

(50) $$R(\lambda) = \sqrt{\frac{2}{\pi}}\; e^{-\frac{x^2}{8}} \cos (x\omega + \pi\omega^2)$$

FORMULES DIVERSES

- A - LES FONCTIONS $\mathcal{L}f$ et $\mathcal{L}g$

Nous désignerons par les notations $\mathcal{L}f$ et $\mathcal{L}g$ les dérivées de ef et e par rapport à l'indice :

(51) $$\mathcal{L}f_\alpha(x) = \frac{\partial}{\partial\alpha} ef_\alpha(x)$$

(52) $$\mathcal{L}g_\lambda(x) = \frac{\partial}{\partial\lambda} eg_\lambda(x)$$

Le choix de la notation vient de ce que certaines primitives qui s'expriment ordinairement par fonctions factorielles présentent des cas d'exception ou les fonctions ci-dessus s'introduisent de la même manière que s'introduit le logarithme dans la primitive de x^K pour K = -1.

On vérifie sans peine que les séries et les intégrales définies qui représentent ef et eg donnent $\mathcal{L}f$ et $\mathcal{L}g$ par dérivation partielle en α et λ ; (pour $\mathcal{L}g$, on se rappellera que l'intégrale (40) n'a plus de sens lorsque $\lambda = \frac{n}{2}$; il faudrait alors dériver deux fois l'intégrale pour obtenir $\mathcal{L}g$).

Ces cas exceptés, on a pour α ou $\lambda > 0$:

(53) $$\mathcal{L}f_\alpha(x) = 4\int_0^\infty e^{-z^2+zx} \cdot z^{2\alpha-1} \cdot \mathcal{L}z\,dz$$

(54) $$\mathcal{L}g_\alpha(x) = \frac{2}{\pi}\int_0^\infty e^{-z^2} z^{+2\lambda-1}\left[2\mathcal{L}z \,.\, \sin(xz+\pi\lambda) + \pi\cos(xz+\pi\lambda)\right] dz$$

De ces expressions on tire des développements asymptotiques par les mêmes méthodes que pour ef et eg. Nous aurons besoin de ceux qui se rapportent aux voisinages de $x = \pm\infty$ ils se calculent sur $\mathcal{L}f$, pour $\alpha > 0$ d'abord, puis pour α négatif par récurrence.

Il existe en effet une relation de récurrence pour $\mathcal{L}f$, que l'on obtient en dérivant (16) par rapport à α :

(55) $$\mathcal{L}f_{\alpha+1}(x) = \frac{x}{2}\mathcal{L}f_{\alpha+\frac{1}{2}}(x) + \alpha\mathcal{L}f_\alpha(x) + ef_\alpha(x)$$

Et de même pour g :

(56) $$\mathcal{L}g_{\lambda+1}(x) + \frac{x}{2}\mathcal{L}g_{\lambda+\frac{1}{2}}(x) + \lambda\mathcal{L}g_{\lambda}(x) + eg_{\lambda}(x) = 0 .$$

On notera, en même temps, que $\mathcal{L}f$ et $\mathcal{L}g$ vérifient les équations différentielles en z ou t :

(57) $$z'' = \frac{x}{2}z' + \alpha z + ef_{\alpha}(x)$$

(58) $$t'' + \frac{x}{2}t' + \lambda t + eg_{\lambda}(x) = 0 .$$

Ces équations s'intègrent facilement puisque (55) et (56) donnent une solution particulière avec second membre et que ef et eg vérifient respectivement les équations sans second membre.

On trouve ainsi

(59) $$z = \mathcal{L}f_{\alpha}(x) + A\ ef_{\alpha}(x) + B\ ef_{\alpha}(-x)$$

(60) $$t = \mathcal{L}g_{\lambda}(x) + C\ eg_{\lambda}(x) + D\ eg_{\lambda}(-x) .$$

Donnons enfin les premiers termes des développements asymptotiques de $\mathcal{L}f$ et $\mathcal{L}g$ au voisinage de

$x \longrightarrow +\infty$

(61) $$\mathcal{L}f_{\alpha}(x) \sim 4\sqrt{\pi}\, e^{\frac{x^2}{4}}\left(\frac{x}{2}\right)^{2\alpha-1}\left\{\mathcal{L}\frac{x}{2}\cdot\left[1+(2\alpha-1)(2\alpha-2)\frac{1}{x^2}+\ldots\right]+\left[(4\alpha-3)\frac{1}{x^2}+\ldots\right]\right\}$$

(62) $$\mathcal{L}g_{\lambda}(x) \sim \frac{2\sqrt{\pi}}{\Gamma(\frac{1}{2}-\lambda)\Gamma(1-\lambda)}\left(\frac{2}{x}\right)^{2\lambda}\left\{\left[\psi\left(\tfrac{1}{2}-\lambda\right)+\psi(1-\lambda)\right]\left[1+2\lambda(2\lambda+1)\frac{1}{x^2}+\ldots\right]+\left[2(4\lambda+1)\frac{1}{x^2}+\ldots\right]-2\mathcal{L}\frac{x}{2}\left[1+2\lambda(2\lambda+1)\frac{1}{x^2}+\ldots\right]\right\}.$$

(62bis) $$\mathcal{L}g_{\frac{k}{2}}(x) \sim 4\frac{(k-1)!}{(-x)^k}\left[1+k(k+1)\frac{1}{x^2}+\ldots\right],$$

$x \longrightarrow -\infty$

(63) $$\mathcal{L}f_{\alpha}(x) \sim 2\left(\frac{1}{x^2}\right)^{\alpha}\Gamma(2\alpha)\left\{2\left[\psi(2\alpha)-2\alpha(2\alpha-1)\psi(2\alpha+2)\frac{1}{x^2}+\ldots\right.\right.$$
$$\left.\left.+\left[2(3-4\lambda)\frac{1}{x^2}+\ldots\right]-\mathcal{L}x^2\left[1-2\alpha(2\alpha+1)\frac{1}{x^2}+\ldots\right]\right\}.$$

(64) $$\mathcal{L}g_{\lambda}(x) \sim \frac{e^{-\frac{x^2}{4}}\Gamma(1-2\lambda)}{\Gamma(1-\lambda)\Gamma(\frac{1}{2}-\lambda)}\left(\frac{1}{x^2}\right)^{\frac{1}{2}-\lambda}\left\{\left[\psi(1-\lambda)+\psi\left(\tfrac{1}{2}-\lambda\right)-2\psi(1-2\lambda)\right]\left[1-(1-2\lambda)(2-2\lambda)\frac{1}{x^2}+\ldots\right]+\left[2(3-4\lambda)\frac{1}{x^2}+\ldots\right]+2\mathcal{L}x^2\left[1-(1-2\lambda)(2-2\lambda)\frac{1}{x^2}+\ldots\right]\right.$$

(64bis) $$\mathcal{L}g_{\lambda}(x) \sim \frac{e^{-\frac{x^2}{4}}}{2\pi}\frac{\Gamma(\lambda)\Gamma(\lambda+\frac{1}{2})}{\Gamma(2\lambda)}\left(\frac{1}{x^2}\right)^{\frac{1}{2}-\lambda}\left\{\left[\psi(\lambda)+\psi(\lambda+\tfrac{1}{2})-2\psi(2\lambda)\right]\left[1-\frac{(1-2\lambda)(2-2\lambda)}{x^2}+\ldots\right]+\left[2(3-4\lambda)\frac{1}{x^2}+\ldots\right]+2\mathcal{L}x^2\left[1-(1-2\lambda)(2-2\lambda)\frac{1}{x^2}+\ldots\right]\right\}$$

- B - LES "FORMES QUADRATIQUES FACTORIELLES".

Nous désignerons ainsi les expressions de la forme

$$\sum A_i \quad ef_{\alpha_i} \; eg_{\lambda_i}$$

dont certaines donnent lieu à des identités remarquables, qui font penser (et ce ne doit pas être une coincidence) aux équations de la physique quantique ("crochets" de Dirac).

Soient u et v deux fonctions factorielles de première espèce, d'indices α et β ; on a

$$u'' = \frac{x}{2}u' + \alpha u$$

$$v'' = \frac{x}{2}v' + \beta v$$

On vérifiera alors sans peine l'identité suivante (en dérivant) :

$$(\alpha - \beta) \int^x uv e^{-\frac{x^2}{4}} \, dx = (u'v - uv') \, e^{-\frac{x^2}{4}} + c^{te}$$

La constante se détermine aisément dans certains cas : si u et v sont des ef, la constante s'annule en prenant $-\infty$ pour borne inférieure d'intégration d'après (43). Si u et v sont toutes deux des cf ou toutes deux des sf, la constante s'annule en prenant O pour borne inférieure d'intégration (en raison de la parité).

L'identité (65) suppose $\alpha \neq \beta$; pour $\beta = \alpha$ on obtient une nouvelle identité par la règle de l'Hospital; en désignant par u_1 la dérivée de u par rapport à l'indice α, on trouve :

(66) $$\int^x u^2 e^{-\frac{x^2}{4}} \, dx = (uu'_1 - u'u_1) \, e^{-\frac{x^2}{4}} + c^{te}$$

En particulier :

(67) $$\int_{-\infty}^x ef_\alpha^2 \, (x) \; e^{-\frac{x^2}{4}} \, dx = (ef_\alpha \mathcal{L} f_{\alpha+\frac{1}{2}} - ef_{\alpha+\frac{1}{2}} \mathcal{L} f_\alpha) \, e^{-\frac{x^2}{4}}$$

Ces derniers résultats expliquent, comme nous l'avions annoncé, la notation $\mathcal{L}$f.

L'EMPLOI DES FONCTIONS FACTORIELLES DANS LES LOIS DE PROBABILITÉ D'ÉTIENNE HALPHEN

(Note de M. P. LARCHER)

Les fonctions factorielles ont été étudiées à l'occasion de recherches sur les lois de probabilités dénommées : "Lois B. Halphen".

Ces lois ont pour densité de probabilité l'expression suivante :

$$f(z) = ke^{-z^2+xz} . z^{2\alpha-1}$$

dépendant des deux paramètres x et α. L'intérêt de ces lois réside dans la grande diversité de formes des courbes de probabilités, suivant les valeurs données à ces paramètres. On obtient en effet des courbes en "J" (infinies à l'origine et décroissantes), des courbes en "S" (infinies à l'origine et présentant un maximum), des courbes de Gauss tronquées et enfin des courbes nulles à l'origine et présentant un maximum, dont les unes sont à l'origine tangentes à l'axe vertical, d'autres à l'axe horizontal, et d'autres enfin ont leurs tangentes à l'origine obliques.

La fonction de répartition pour ces lois est la suivante :

$$F(z) = K \int_0^z e^{-u^2+xu} . u^{2\alpha-1} . du$$

et comme on doit avoir $F(\infty) = 1$, la constante K a pour valeur :

$$K = \frac{1}{\int_0^\infty e^{-u^2+xu} . u^{2\alpha-1} . du} = \frac{2}{ef_\alpha(x)} \text{ d'après (33)}$$

On voit immédiatement que les différents moments ont pour expression:

$$m_k = \frac{ef_{\alpha+k/2}(x)}{e f_\alpha(x)} .$$

En particulier, la moyenne désignée par la notation $k f_\alpha(x)$ est :

$$k f_\alpha(x) = \frac{e f_{\alpha+1/2}(x)}{e f_\alpha(x)} .$$

C'est la dérivée logarithmique de $e f_\alpha(x)$ par rapport à x.

Sa dérivée par rapport à x est :

$$\frac{\partial\, kf_\alpha(x)}{\partial x} = \frac{ef_{\alpha+1}(x)\, ef_\alpha(x) - ef^2_{\alpha+1/2}(x)}{ef^2_\alpha(x)} = \frac{ef_{\alpha+1}(x)}{ef_\alpha(x)} - \left[\frac{ef_{\alpha+\frac{1}{2}}(x)}{ef_\alpha(x)}\right]^2$$

$$\frac{\partial\, kf_\alpha(x)}{\partial x} = m_2 - m_1^2 = \sigma^2 .$$

La dérivée de la moyenne par rapport à x est égale à la variance.

Dérivons maintenant $k\, f_\alpha(x)$ par rapport à α :

$$\frac{\partial\, kf_\alpha(x)}{\partial\alpha} = \frac{\mathcal{L} f_{\alpha+1/2}\, ef_\alpha - ef_{\alpha+1/2}\, \mathcal{L} f_\alpha}{ef^2_\alpha} .$$

La relation (67) montre que cette dérivée est toujours positive. La moyenne est donc une fonction croissante de α.

La moyenne géométrique de ces lois s'exprime de la façon suivante, en la désignant par g :

$$\text{Log } g = \frac{2\int_0^\infty e^{-u^2+xu}\, u^{2\alpha-1}\, \text{Log } u\, du}{ef_\alpha(x)} = \frac{1}{2}\, \frac{\mathcal{L} f_\alpha(x)}{ef_\alpha(x)}$$

d'après (53). Son logarithme naturel est égal à la demi-dérivee logarithmique de $ef_\alpha(x)$ par rapport à α.

Moyenne, écart quadratique moyen et moyenne géométrique fournissent pour les lois B l'ajustement du maximum de densité ("maximum likelihood").

LA NOTION DE VRAISEMBLANCE

par

Étienne HALPHEN

PRÉFACE

par

G. MORLAT

Etienne HALPHEN avait écrit ce mémoire en 1947. Des vicissitudes diverses empêchèrent à l'époque sa publication. Certes, il ne faut pas s'attendre à y trouver exclusivement des idées neuves. Mais, dans ce problème des fondements de la probabilité, qui est encore loin d'être clos, de vieux points de vue peuvent être rajeunis et nous apporter des clartés nouvelles. Le livre publié récemment par J.L. Savage en porte témoignage(1) Nous croyons que le présent mémoire d'Etienne Halphen en sera un autre exemple.

D'aucuns croirons sans doute reconnaitre ici les théories de H. Jeffreys, d'autres celles de J. M. Keynes, d'autres encore... Mais pour couper court à ces suppositions, qu'il me soit permis, sans déroger à la grande admiration que j'éprouve pour Etienne Halphen, de faire pour lui un aveu : il n'avait pas lu tous ces excellents auteurs, et sans doute il en eût été incapable. En mathématiques, il avait lu bien peu de choses. Il préférait réinventer ce dont il avait besoin. Ce n'est point là sans doute une attitude qui peut être conseillée à beaucoup. Il y faut une intelligence exceptionnelle - et probablement du génie. Alors, pour qui sait réinventer à sa façon ce qui a été dit avant lui, et même aller un peu plus loin, on peut penser que la lecture est superflue, sinon inhibitrice. Etienne Halphen était dans ce cas.

On croira que j'exagère ? - Pensera-t-on qu'il aurait imaginé la belle théorie des fonctions factorielles (2), s'il avait commencé par étudier les résultats déjà connus concernant ces fonctions - sous le nom de fonctions de Hermite ?

Mais peut-être le présent ouvrage convaincra le lecteur de ce que j'avance. Si le point de vue d'où part l'auteur n'est pas nouveau, les développements qu'il donne contiennent maints aspects originaux. Soulignons par exemple l'accent mis sur la différence de nature entre vraisemblances

---

(1) The Foundations of statistics. Wiley. New York. 1954.

(2) Mémoire qui parait dans ce même fascicule.

et probabilités - le rôle dévolu à la "caractérielle estimatrice" dans le problème de l'estimation des paramètres (1) - et encore les indications données à la fin du chapitre IV sur le problème des tests. Sur ce dernier point en vérité, il convient de noter qu'Etienne HALPHEN ne croyait pas que l'interprétation qu'il avait donnée à la notion de test fût définitive. Simplement, elle lui paraissait mieux correspondre au désir intuitif du "client" que la théorie générale des tests de Neyman et Pearson.

*

* *

Je ne sais si c'est à propos de cette Théorie générale des tests, ou bien à propos de la Théorie des Intervalles de confiance de Neyman - d'ailleurs les deux se ressemblent si fort - qu'Etienne Halphen un jour déclara : "Mais ce n'est pas de la statistique, c'est du contrôle des fabrications.". Même si on ne le suit pas dans cette dénomination qui risque de paraitre péjorative, il est permis de regretter que les traités de statistique ne disent pas plus nettement la distinction que sépare les deux classes de techniques: celles qui visent à fonder une décision sur la plus grande vraisemblance - au sens de Halphen - et celles qui visent à réaliser dans nos décisions successives une certaine fréquence de réussites. Ces deux classes de techniques correspondent à des conceptions - ou plutôt à des applications - radicalement opposées du calcul des probabilités : grosso modo, le calcul des probabilités schématise pour les premières les "degrés de croyance rationnelle" de Keynes, les "vraisemblances" de Halphen (intuitionnistes) - pour les secondes, la limite des fréquences empiriques et elle seule (fréquentistes) (2).

Il faut s'attendre à ce que ces techniques ne traitent pas les mêmes problèmes. Jamais les intervalles de confiance ne donneront une idée quelconque des valeurs plus ou moins vraisemblables du paramètre auquel ils s'appliquent - jamais l'estimatrice de Halphen ne se prêtera à un calcul précis des fréquences de réussite lorsqu'on y assigne au paramètre un certain intervalle.

Mais dans beaucoup de problèmes qui se posent à l'ingénieur, il s'agit de prendre une décision, fondée sur une valeur d'un paramètre incertain ; c'est-à-dire que seul un procédé d'estimation ponctuel satisfait le client. Ce n'est guère qu'à propos de contrôle des fabrications et de problèmes similaires qu'une décision peut consister dans un intervalle. Cela explique la boutade un peu forte d'Etienne Halphen.

Il faut ajouter que si la plupart des auteurs de traités ont manqué, comme je le déplore, à faire la distinction, c'est pour la raison fort simple qu'ils ont pris parti. Cela signifie qu'ils ont adopté une des catégories de techniques et nient la légitimité de l'autre. C'est fort regrettable.

On sait que les savants anglo-américains dominent de loin la statistique moderne. La majorité d'entre eux s'est ralliée au point de vue fréquentiste. Une certaine justification peut être trouvée dans la prolifération et l'efficacité des techniques qui s'y rattachent. Mais on sait que beaucoup

---

(1) Notons qu'Etienne Halphen rejetait formellement l'emploi des probabilités a priori tel que le recommande H. Jeffreys...

(2) Nous conservons ici, pour désigner la controverse fondamentale de la Théorie de la Probabilité, les termes "intuitionniste" et "fréquentiste", qui avaient l'agrément d'E. Halphen. Dans son livre déjà cité, J.L. Savage désigne les adversaires par les termes : "personnalistes" et "objectivistes" ; il leur adjoint d'ailleurs une troisième catégorie comprenant quelques logiciens exigeants.... Il précise utilement qu'une classification aussi succincte est propre à mettre en furie tous les spécialistes de la question.

de ces techniques sont aussi bien cohérentes avec l'autre point de vue (c'est ce que J.L. Savage s'est attaché à montrer) . Et pour d'autres (Intervalles de confiance notamment) Etienne Halphen remarque que leur emploi, dans beaucoup de problèmes, ne répond nullement au désir du client - remarque que nous avons retrouvée à la fin du livre de J.L. Savage.

*

* *

Disons, pour achever cette préface, un mot encore sur le rôle qu'Etienne Halphen assignait à la Statistique. Dans le mémoire qu'on va lire, apparait déjà un souci très fort : celui de soumettre les techniques statistiques aux préoccupations du "client" (ingénieur, entrepreneur ou économiste, etc...). C'est bien ce souci qui conduit à rejeter certaines techniques trop préoccupées de leur propre logique.

Après des années de réflexion, Etienne Halphen allait très loin dans cette voie, puisqu'il écrivait peu avant sa mort ces phrases dans lesquelles il ne faut voir aucune modestie :

" La statistique ne doit pas résoudre les problèmes ; elle doit contribuer à former le jugement de ceux qui ont à les résoudre ; mais il importe qu'ils ne les résolvent qu'après avoir "oublié" la statistique.
J'ai naguère commis une erreur en déclarant la statistique "science des décisions" ; elle est seulement "école de formation du jugement" ; celui qui prend une décision ne doit pas être statisticien : il doit croire un peu à la statistique, mais pas trop.
Quant au statisticien, il doit modérer la foi excessive des uns, mais aussi combattre le scepticisme systématique des autres."

LA NOTION DE VRAISEMBLANCE

ESSAI SUR LES FONDEMENTS DU CALCUL DES PROBABILITÉS ET DE LA STATISTIQUE MATHÉMATIQUE

INTRODUCTION

Les réflexions qui suivent ont eu comme point de départ certaines difficultés rencontrées dans la pratique de la statistique mathématique. Ayant eu professionnellement à résoudre des problèmes d'estimation sur des données très pauvres, nous nous sommes rendu compte que les auteurs des méthodes modernes d'estimation ne semblaient pas avoir une idée bien précise du but à atteindre. Si, à la différence des auteurs anciens, ils ont la préoccupation de réaliser un optimum, ils sont fort embarrassés lorsqu'il s'agit de définir en quoi celui-ci doit consister ; en fait, ils se donnent cet optimum de façon assez arbitraire, selon leurs goûts ou préférences philosophiques. C'est oublier que le statisticien ne travaille pas pour lui-même, mais qu'il est au service du praticien qui le consulte ; industriel, économiste, biologiste... ; c'est à ce praticien de fixer le but à atteindre, de dire ce qu'il a besoin de rendre minimum.

Pour satisfaire ce praticien, le statisticien a à sa disposition, outre diverses données expérimentales, sa propre intelligence : celle-ci lui fournit, non seulement des "techniques" de calcul, mais encore certaines intuitions, certaines "vraisemblances", dont la nature et l'origine peuvent être obscures, mais dont personne ne songe à nier la réalité. Seulement on considère habituellement que ces "vraisemblances" ne doivent intervenir qu'après la technique, pour corriger celle-ci, et non point avant : mais c'est là toute la question.

Nous sommes arrivé pour notre compte à la conviction que, tout au contraire, les vraisemblances doivent intervenir (et de fait, qu'on le veuille ou non, interviennent) avant tout calcul : leur rôle dépasse de loin une simple correction de bon sens à ce que le calcul peut avoir de trop rigide ; leur rôle est de fournir la "matière première" que le calcul aura pour fonction de dégrossir.

C'est ce que nous pensons pouvoir montrer, en étudiant dans un premier chapitre quelques exemples de difficultés qui nous semblent insolubles autrement. De toute manière, nous demandons que la question soit examinée, non selon telle ou telle préférence philosophique à priori, mais selon cette idée directrice : la meilleure méthode statistique sera celle, quelle qu'elle soit, qui permettra au statisticien (homme vivant) de répondre au mieux aux exigences de celui qui le consulte.

Faisons observer, en outre, pour ceux qu'effrayerait la part que nous faisons à l'intuition, qu'intuitif n'est pas synonyme d'arbitraire : et qu'au reste assez nombreux sont les cas où convergent les intuitions de tous les hommes, pour qu'habituellement l'usage de la statistique mathématique ne prête pas à contestation.

Après avoir ainsi indiqué dans un premier chapitre comment nous entendons poser le problème, et pourquoi, nous proposerons notre solution: le chapitre II sera consacré à l'étude des notions psychologiques indispensables ; le chapitre III expliquera ce que nous appelons "vraisemblances" et montrera comment cette notion déborde celle de probabilités et la fonde : le chapitre IV fera la jonction entre ces notions théoriques et la science pratique du statisticien.

Nous verrons alors que presque toutes les méthodes d'estimation qui ont été proposées sont bonnes, mais chacune dans un cas déterminé ; il n'y a pas une méthode meilleure que les autres ; il y a des méthodes dont chacune est la meilleure dans son propre champ d'application, car chacune répond à un type de question, différent de l'une à l'autre.

Nous verrons aussi que les différences numériques entre les résultats de ces méthodes s'atténuent très vite lorsque le nombre des épreuves augmente, si bien que les divers procédés d'estimation sont en pratique presque équivalents. Reste que leurs différences théoriques présentent un grand intérêt spéculatif.

CHAPITRE PREMIER

REMARQUES SUR LE PROBLÉME DE L'ESTIMATION[1]

Les méthodes anciennes d'estimation comportaient assurément une grande part d'arbitraire : dans l'estimation d'une loi de Gauss, pourquoi utiliser la moyenne plutôt que la médiane, l'écart-type plutôt que l'écart-moyen, ou l'interquartile ? on ne le savait pas au juste...

Depuis les travaux de Fisher, on se préoccupe de réduire cette marge d'arbitraire et de trouver des principes d'estimation doués d'une justification rationnelle. On savait déjà que toute méthode d'estimation n'est pas bonne, que, par exemple, on ne peut estimer une loi de Cauchy par la moyenne : Fisher a donné quelques règles simples pour distinguer les méthodes en deux classes ; les méthodes efficaces et les méthodes inefficaces. Mais ce n'est pas suffisant. Parmi les méthodes efficaces, toutes ne se valent pas : quelle sera la "meilleure" ? Fisher a donné une solution du problème dans le cas particulier d'observations très nombreuses ; mais dans le cas général, la question reste ouverte.

1. COMMENT SE POSE LE PROBLÈME

Pour rechercher la "meilleure méthode d'estimation, il convient de se poser deux questions :

1°) Quel est le but que l'on se propose ?

2°) Quels sont les moyens à notre disposition pour y parvenir ?

Or, une chose est frappante lorsqu'on lit les divers travaux modernes sur l'estimation : chaque auteur commence par choisir une définition à lui du but à atteindre ; ce but, qui devrait nous être imposé par la nature même des choses et le rôle subordonné de la science statistique par rapport aux autres sciences ou aux arts de l'ingénieur, ce but, chaque statisticien s'arroge le droit de le fixer à sa convenance.

En voulez-vous un exemple frappant ? Examinons comment Neyman parvient à sa théorie des "unbiassed sets of confidence intervals".

Il commence par se demander si l'on pourrait construire un ensemble d'intervalles répondant à un certain ensemble de deux conditions A et B : il s'agit de conditions qui, si on pouvait en effet les remplir, seraient fort satisfaisantes. Mais Neyman constate que la condition B ne peut en

(1) L'essentiel de ce chapitre avait été exposé au Colloque de Calcul des Probabilités, tenu à Lyon en 1947 sous l'égide du C.N.R.S.

général être vérifiée. Alors, il lui substitue une condition B' qui lui ressemble beaucoup du point de vue mathématique, et comme les conditions A et B' peuvent être simultanément satisfaites, le problème parait résolu. Mais, si l'on y regarde de plus près, on s'aperçoit que la ressemblance entre B et B' est surtout formelle, et que la dernière n'a plus qu'une signification pratique assez arbitraire : elle ne répond plus à un besoin profond de la statistique, elle ne s'impose aucunement.

Que l'on me permette une comparaison familière ; supposons que j'aie besoin d'un ciment répondant à une certaine exigence très précise ; le commerçant à que je m'adresse n'a pas l'article qu'il me faudrait, mais, abusant de mon incompétence, il m'en glisse un autre : "Prenez cela, c'est absolument ce qu'il vous faut". Evidemment, il me vend celui de ses produits qui se rapproche le plus de mes désidérata ; mais enfin ce commerçant est-il parfaitement honnête ? Et sera-ce pour lui une justification suffisante de dire ; 'Je vous ai vendu ce ciment parce que je me refuse à tenir la qualité que vous me demandez et qui ne me plait pas....". ?

Car voici le point où nous aurons à en venir : si Neyman et d'autres se donnent telle ou telle définition à priori des exigences de l'estimation statistique, tout se passe comme s'ils avaient décidé à l'avance d'éliminer certaines méthodes qui ne répondent pas à leur conception de la science. Quelles méthodes ? Celles qui font une place trop grande à l'intuition, au "flair" du statisticien. Ceci nous mène à l'étude du second point.

QUELS SONT LES MOYENS A LA DISPOSITION DU STATISTICIEN ?

On peut les ranger sous trois rubriques : les sciences expérimentales dans leur état actuel ; les "échantillons" soumis au statisticien ; l'intelligence au sens le plus large, sans en rien exclure ; donc une intelligence concrète, celle d'un homme déterminé pris avec toutes ses facultés vivantes. Tous ces moyens peuvent jouer un rôle dans l'estimation ; tous peuvent contribuer à la rendre plus précise, tous aussi peuvent y introduire leur part d'erreur.

Parmi ces erreurs, celles qui viennent de la mauvaise qualité des échantillons sont étudiées à part, sous le nom précisément d'erreurs d'observation. Celles qui sont dues à l'état imparfait de la science posent le problème dit des "tests". Celles qui proviennent enfin de l'intelligence du statisticien sont attribuables partie aux méthodes statistiques dont il fait usage, partie, à sa justesse personnelle d'appréciation.

La tendance actuellement dominante consiste à raisonner ainsi ; appelons α les erreurs imputables à la méthode, β celles qu'introduit l'appréciation personnelle ; pour rendre minimum la somme $\alpha + \beta$, il faut minimiser séparément α et β. Conclusion : parmi les conditions que doit remplir la "meilleure" estimation, se trouve l'élimination aussi complète que possible de l'intuition personnelle ; ensuite, on cherchera à rendre α aussi petit que possible.

Ce raisonnement (implicite) serait parfaitement juste si α et β étaient indépendants ; seulement est-ce le cas ? Nous pensons que non (1). Et pour preuve, nous en donnons les efforts extraordinaires accomplis depuis un quart de siècle pour fonder en dehors de tout recours à l'appréciation personnelle une méthode "objective" d'estimation, efforts qui tous viennent

(1) Nous pensons que tout se passe à peu près comme si le produit $\alpha\beta$ était constant (cf. la relation de Heisenberg), en sorte qu'à vouloir annuler β on rendrait α infini.

se heurter à des difficultés étranges ; chaque tentative nouvelle commence par une marche victorieuse ... pour venir s'empêtrer dans des "à peu près" ou des "en admettant que" qui montrent finalement que l'intuition ne se laisse pas éliminer comme on l'avait cru, ou qu'à vouloir l'éliminer on perd tout contact avec le réel : on résout alors de beaux problèmes qui ne sont pas ceux que nous posent les hommes qui nous consultent (biologistes, ingénieurs ...) ; comme le marchand de ciment malhonnête, nous ne fournissons pas à notre client l'article qu'il nous demande.

Je ne veux pas dire que les théoriciens de l'estimation veulent explicitement éliminer le facteur intuition : mais cela leur semble aller tellement de soi que, dans la façon même dont ils posent le problème, tout recours à ce facteur est déjà éliminé.

Or, il ne s'agit pas de contester les risques d'erreur qu'introduit l'intuition personnelle du statisticien ; mais pourquoi, d'abord, voudrait-on que la statistique soit de ces sciences faciles où le premier venu peut réussir à condition de respecter un mode d'emploi quasi mécanique ? On nous dira que la statistique est une science et ne doit pas reposer sur de l'arbitraire ? Il s'agit de savoir si intuitif et arbitraire sont synonymes. Surtout, il s'agit de savoir si, le but du statisticien étant de répondre au mieux à des questions qu'un "client" lui a posées, la meilleure méthode ne sera pas celle qui donnera en effet la meilleure réponse - que ce résultat soit atteint avec ou sans recours à l'intuition.

De deux statisticiens, dont l'un fait un large appel à son "flair" et trouve des réponses précises, et dont l'autre, à l'aide de méthodes plus standardisées, répond moins bien : lequel est le meilleur ?

Le problème de la meilleure méthode d'estimation ne peut être séparé de celui du meilleur statisticien ; et ce dernier doit être jugé en fonction des exigences de celui qui le consulte ; ce n'est pas au statisticien de fixer ces exigences : il a seulement à les bien comprendre, parfois à demi-mot, ce qui demande de sa part un minimum de compétence dans la matière en question ; on oublie cela trop souvent.

2. QUELQUES EXEMPLES

Nous sommes donc arrivés à ce résultat : le meilleur statisticien, c'est celui qui satisfait au mieux aux exigences de "client" ; si, pour ce faire, il doit réduire le recours à l'intuition, qu'il le fasse ; sinon, qu'il ne le fasse pas. Voici quelques exemples de problèmes qui feront bien ressortir les difficultés auxquelles achoppent les théories actuelles de l'estimation.

(a) Mon client est la Société Nationale "Electricité de France", qui me demande de prévoir au mieux la puissance H à donner aux usines hydro-électriques en 1951 pour faire face aux besoins de la consommation (1). Si je donne une évaluation trop faible, il y aura des coupures de courant, d'où inconvénients sociaux et pertes économiques ; si je donne un chiffre trop fort, les dépenses d'équipement seront inutilement élevées. L'évaluation exacte réaliserait un optimum économique.

Pratiquement, je ne pourrai faire qu'une estimation approchée de H. La valeur exacte H_0 réalisant un optimum, on peut admettre qu'au voisinage de H_0 la perte économique résultant d'une estimation inexacte, est approximativement proportionnelle à $(H - H_0)^2$. C'est donc cette quantité que je dois chercher à rendre minimum.

(1) Rappelons que ce mémoire a été écrit en 1947.

Aitken a étudié ce type de problème. Mais à la condition précédente, il a ajouté la condition : E (H) = H_0 , exigée par ce que Fisher appelle "unbiassedness". Mais c'est là un nouvel exemple de condition imposée sans nécessité par le statisticien. Si cette condition résulte de la précédente, très bien : mais en ce cas, inutile de l'écrire spécialement ; or, il reste à démontrer qu'il en est ainsi, En attendant, pourquoi nous en occuper ? Cela ne peut que rendre moins satisfaisant le minimum de $(H - H_0)^2$

Voici en réalité d'où est venue cette condition.

Lorsque la statistique n'avait pas encore entrepris une étude systématique du problème de l'estimation, presque toutes les estimations consistaient en des moyennes. Lorsqu'on passait d'un échantillon de 10 individus à un échantillon de 100 individus, on faisait tout naturellement la moyenne des 10 moyennes précédentes. On se rendit compte alors que l'opération n'était pas toujours correcte, et exigeait une certaine correction dite "d'erreur systématique".

Mais rien ne nous oblige à passer des petits échantillons au grand en faisant une moyenne ; en réalité, nous n'avons aucune raison de faire ainsi, mais nous devons traiter le grand échantillon pour lui - même (1) : un grand échantillon n'est pas une itération de petits échantillons.

(b) Mon "client" m'a posé le problème suivant :

"Je sais qu'une variable X suit une loi de Gauss d'écart-type connu : pour la moyenne, j'ai le choix seulement entre deux valeurs a, b, $(a < b)$. Je ne connais qu'une seule valeur de X, x : laquelle des deux lois dois-je adopter ?"

Reconnaissons que la donnée d'une seule valeur de X est un cas très extraordinaire ; ordinairement on refusera de se prononcer dans ces conditions. Mais :

D'abord, mon client peut me demander absolument de répondre, sachant certes que ma réponse aura peu de valeur ; mais, obligé pour son compte de faire un choix, il désire faire le moins mauvais des deux,

ensuite, ce problème ne diffère pas, qualitativement, de tout autre problème d'estimation : on me demande de faire au mieux, on n'exige pas que ce mieux soit très bon. Si réellement la statistique est scientifique, elle doit permettre de fournir, non pas toujours une bonne estimation, mais toujours la meilleure ou si l'on veut la moins mauvaise.

enfin, si j'envisage ce cas limite, c'est simplement pour ne pas compliquer le problème essentiel de l'estimation par des problèmes parasites.

Maintenant je pense que tous les statisticiens seront d'accord dans la solution du problème précédent :

Ils opteront pour a ou pour b selon que X sera inférieur ou supérieur à $\frac{a + b}{2}$.

Or, si nous examinons les présupposés implicites de cette solution, voici ce que nous constatons :

---

(1) En théorie, bien sûr. En pratique, il sera souvent commode de procéder par moyennes, et il n'en résultera pas d'inconvénients sérieux. Mais nous parlons ici de principes.

Considérons deux statisticiens S_1 et S_2 à qui est posé le problème précédent, et, dans la suite, d'autres problèmes semblables. Si S_1 applique la solution ci-dessus, mais que S_2 , doué d'un certain "flair", favorise dans une certaine mesure soit une loi soit l'autre selon certaines intuitions (impossibles du reste à formuler en termes mathématiques) : alors, si les intuitions de S_2 sont justes, il tombera dans l'ensemble un peu plus souvent juste que S_1 .

Qu'est-ce que cela signifie ? Que la solution dont nous avons parlé n'est la meilleure que pour autant qu'on tient (indépendamment de l'expérience X), a et b pour également vraisemblables ; ce n'est donc pas à proprement parler un état négatif d'ignorance qui nous a fait procéder comme nous l'avons fait mais plutôt un état positif d'égale vraisemblance.

Notre méthode n'est la meilleure que pour un statisticien qui tient ainsi a et b pour également vraisemblables. Cela montre que le problème de la meilleure méthode statistique est un problème postérieur à celui du meilleur statisticien.

Pour comparer deux méthodes statistiques, il faut se demander laquelle donnera les meilleurs résultats, à supposer que toutes deux soient appliquées par le même statisticien, et à supposer que celui-ci soit un bon statisticien.

Un bon statisticien, c'est d'abord celui qui a un bon "flair" ; une bonne méthode, c'est celle qui réussit la mieux lorsque maniée par un bon statisticien.

Nous reprochons donc aux méthodes de Fisher et de Neyman d'intervertir l'ordre des choses ; de vouloir définir la meilleure méthode avant de définir le meilleur statisticien.

L'exemple suivant rendra ces idées encore plus claires :

(c) On me demande d'estimer la corrélation entre deux variables X et Y dont on me donne une série de valeurs expérimentales. La valeur calculée brutalement du coefficient de corrélation se trouve être 0,8. Mais incidemment, on me signale que X désigne la température moyenne quotidienne à la station de Paris Parc-St-Maur, tandis que Y est le dernier chiffre du numéro matricule de la première auto croisée par mon client en sortant de chez lui chaque matin.

Mon intuition de statisticien va-t-elle entrer en jeu après ou avant d'avoir trouvé 0,8 pour coefficient de corrélation entre X et Y ?

J'ai pris ici un cas limite. Mais il en est quantité d'autres où, sans en être certain, je tiens la corrélations entre X et Y pour vraisemblablement faible plutôt que forte, ou bien le contraire : ainsi, il est vraisemblable qu'il y a une correlation nettement positive entre l'élégance des costumes et les revenus des passants que je croise dans la rue.

Le meilleur statisticien, ici encore, sera celui qui, ayant d'abord fait preuve d'un bon "flair", achèvera le calcul par la meilleure méthode.

(d) Le problème des ajustements inadéquats.

Voici un nouveau type de difficultés. Les méthodes d'estimation de Fisher ou de Neyman supposent essentiellement que la loi de probabilité cherchée, inconnue en elle-même, appartient à coup sûr à une certaine classe bien déterminée : loi de Gauss, loi de Galton...

Or, ceci n'est presque jamais le cas. Habituellement, le statisticien sait parfaitement que, s'il est obligé de faire une "hypothèse de classe", de ne chercher la loi inconnue que par précision sur les constantes d'un type analytique déterminé, ce n'est que pour la commodité : la classe envisagée ne renferme presque sûrement pas la loi véritable de façon exacte, mais seulement en première approximation. Dans ces conditions, la meilleure méthode d'estimation n'est pas nécessairement celle qui serait la meilleure si l'hypothèse de classe était parfaitement correcte.

Exemple :

J'utilise en première approximation la loi de Galton pour représenter une variable X dont la distribution ne m'intéresse que pour les grandes valeurs. En ce cas, l'estimation par la moyenne et la variance du logarithme donnerait trop de poids aux petites valeurs de X, qui ne m'intéressent pas et, puisque la loi de Galton n'est probablement qu'une loi approchée, je préfèrerai utiliser la moyenne et la variance de X qui rendront mon ajustement plus correct dans la région qui m'intéresse. Nous avons donné à ce problème le nom de : "problème des ajustements inadéquats".

C'est le problème qui se pose pratiquement presque toujours au statisticien.

(e) Une conception traditionnelle mais inexacte de l'échantillonnage.

Nous avons rencontré dans ce qui précède une question importante et délicate.

Ordinairement, pour caractériser la meilleure méthode, les théoriciens supposent l'application réitérée de celle-ci à divers échantillons extraits d'une même population. Cela est inadmissible, pour la raison suivante :

ou bien je saurai que les divers échantillons étudiés proviennent d'une même population : en ce cas je serais impardonnable de n'en pas tenir compte, et dès lors je n'aurai pas à opérer sur une suite d'échantillons juxtaposés, mais sur un unique échantillon qui ira croissant avec le temps.

Ou bien j'ignore que ces échantillons proviennent d'une même population. En ce cas, s'il existe une méthode réalisant en fait l'optimum absolu par rapport à chaque population, elle réalisera aussi l'optimum sur l'ensemble des diverses populations que j'ai à étudier ; mais, s'il n'existe pas de telle méthode, j'introduis en la recherchant une exigence inadmissible : je lâche la proie pour l'ombre ;

C'est toujours l'erreur dénoncée plus haut : on introduit une condition parasite B qui correspond seulement à un désir a priori du théoricien et de ce fait on remplit moins bien la condition A posée par le client - laquelle seule devrait compter.

3. QUELQUES CONCLUSIONS

Non pas une conclusion (ce serait prématuré), mais quelques conclusions partielles.

Il est acquis que la statistique n'est scientifique que dans la mesure où elle se tient en liaison avec le calcul des probabilités. Mais elle ne se réduit pas à ce dernier, et ne s'y réduira jamais, pour la raison fort sim-

ple que le calcul des probabilités est une construction idéale vide de tout contenu réel : un mathématicien peut faire de la géométrie Euclidienne, Riemannienne ,Hilbertienne... à priori : il faut un physicien pour savoir laquelle de ces géométries correspond à la réalité. On peut donc dire que réciproquement le calcul des probabilités ne sera fécond que s'il se tient en liaison avec la statistique ; sinon il résoudra de beaux problèmes qu'il se sera donnés à lui-même pour son amusement, mais qui n'ont aucun intérêt.

Il est acquis que la statistique doit se préoccuper de donner à ses méthodes un certain optimum d'efficacité. Mais cet optimum ne doit pas être décidé à priori en vertu d'une préférence philosophique : il doit correspondre à la fonction même de la statistique, qui est de satisfaire au mieux à certains besoins des praticiens.

Je pense avoir montré que dans la recherche de cet optimum, on ne doit pas perdre de vue que les besoins à satisfaire sont ceux d'un homme vivant qui s'est proposé un certain but (dont nous n'avons pas à discuter), et qui nous demande le meilleur moyen d'y parvenir.

On ne doit pas oublier non plus que le statisticien lui-même est un homme vivant, en sorte qu'on ne peut pas séparer (du moins pas à priori) le problème de la meilleure technique statistique de celui du meilleur statisticien. Personnellement, nous considérons que c'est la question du meilleur statisticien qui se pose d'abord, celle de la meilleure technique venant ensuite et dépendant de la première. La meilleure méthode est celle qui, appliquée par le meilleur statisticien, donne les meilleurs résultats. On n'hésitera donc pas à y faire figurer, outre les caractéristiques de l'objet étudié, celles qui seront convenables pour tenir compte du sujet : nous avons déjà dit que subjectif n'était pas synonyme d'arbitraire : la science n'a pas besoin pour être science d'être mécanisée.

Nous avons donné à l'appui de notre opinion certains arguments qui ne sont peut être pas décisifs : mais venez, mettez-vous à la tâche, essayez de fournir à un praticien la solution d'un problème qu'il vous aura posé, en ne vous préoccupant de rien d'autre que de répondre le mieux possible à sa demande : cela vous convaincra assez vite.

Et cela vous convaincra aussi, s'il en était besoin, que les principes de l'estimation ne sont pas arbitraires, ne résultent pas d'un choix gratuit. Nous ne sommes pas actuellement en état de toujours expliquer de façon parfaitement satisfaisante pourquoi la loi de Gauss est généralement plus convenable que la loi de Laplace ; mais, si nous ignorons le pourquoi, nous sommes certains du fait, et le statisticien qui ferait habituellement usage de la loi de Laplace se tromperait à la longue beaucoup plus souvent que celui qui utilise celle de Gauss.

Il est parfaitement vrai que le statisticien a souvent à faire des choix. et que ces choix ne résultent pas d'une pure déduction mathématique : il ne s'ensuit pas pour autant qu'ils soient arbitraires ; ils se fondent sur un instinct profond, une sorte de "flair" professionnel, qui est une véritable intuition que tel choix est bon, tel autre mauvais.

*

* *

Je vous demande la permission de terminer ces réflexions par une remarque d'ordre philosophique.

Nous nous trouvons actuellement devant deux conceptions de la science :

1°) Une conception idéaliste dûe à Descartes et Kant, pour laquelle la mathématique est l'idéal scientifique universel, de telle sorte qu'une science est jugée d'autant plus "science" qu'elle est plus mathématique, c'est-à-dire qu'elle n'opère que sur des éléments plus ou moins inhérents à l'esprit humain.

On concède bien la nécessité de recourir à l'expérience pour commencer : mais on cherchera à réduire au maximum ce facteur jugé "irrationnel". La science parfaite sera celle qui parviendra, à la limite, à tout "déduire" : c'est cet idéal qu'un Hegel avait cru réaliser.

2°) Une conception réaliste : celle-ci admet qu'il existe deux sortes d'objets de connaissance : a) des objets idéaux, b) des objets concrets ; et qu'a chaque sorte d'objets correspond un mode de connaissance qui lui est propre et donc un type adéquat de science :

a) aux objets idéaux correspond une connaissance immédiate, et une science du type mathématique procédant par voie purement logique ;

b) aux objets concrets, correspond une connaissance médiate, donc une science expérimentale qui suppose dans l'homme une faculté d'appréhension de ces objets extérieurs à l'homme : une intuition intellectuelle objective.

Kant a nié l'existence de cette intuition, mais sans justifier sa négation par la moindre preuve. Et depuis Kant beaucoup de philosophes font comme lui, consciemment ou non, et des savants leur emboitent le pas parce qu'ils font passer la philosophie avant la science. Mais un statisticien de métier, celui qui travaille "sur le chantier" et qui fait de la science d'abord, sans idées préconçues, sait parfaitement que Kant est dans l'erreur, et qu'il existe une intuition intellectuelle objective : le statisticien de métier se trouve devant la philosophie de Kant comme Diogène d'Apollonie devant les Eléates : on nous démontre, par des raisonnements abstraits, que le mouvement ne peut pas exister - mais cela ne nous empêche pas de marcher.

L'intuition du réel est un faculté mystèrieuse ? Oui. Mais le problème est seulement de savoir si elle existe. On nous concède qu'elle existe, mais qu'il faut arriver progressivement à en réduire l'usage comme "non scientifique" ? Mais :

Si le réel existe (et nous l'admettons, je pense ?) nous ne le connaissons que par intuition (cela est si vrai que Kant, niant l'intuition intellectuelle objective, nie de ce fait que nous ayons du réel une vraie connaissance). Dès lors, réduire le recours à l'intuition sera ipso facto réduire le réalisme de notre science : pourquoi cette castration volontaire ? Nos facultés ne sont-elles pas déjà suffisamment limitées par nature, que nous voulions les réduire encore de nos propres mains ?

Et pourquoi vouloir qu'il n'existe qu'un seul type de science ? S'il existe plusieurs types d'objets de connaissance, ne doit-il pas leur correspondre plusieurs types de sciences, chacune appropriée à son objet ? Une science de type idéaliste (la mathématique) pour les objets idéaux ; une science réaliste pour les objets concrets .- Ce qui n'empêchera pas la science réaliste d'utiliser aussi les résultats de la science mathématique : les sciences ne peuvent se contredire, mais doivent s'entr'aider ; nous prostestons seulement contre l'ambition d'une science à vouloir absorber les autres : à chacune son rôle.

Le vrai problème :

Non pas chasser l'intuition,

Mais la contrôler, la "domestiquer" - et dans la mesure du possible la rendre plus claire : c'est cela l'idéal scientifique des sciences du concret. Pourquoi vouloir imposer à ces dernières un idéal qui n'est pas le leur ?

La véritable intelligence consiste-t-elle à tout réduire aux dimensions standard d'un lit de Procuste, ou bien à traiter chaque chose de la manière qui lui convient ?

CHAPITRE II

LES GRANDEURS PSYCHOLOGIQUES

REMARQUES PRÉLIMINAIRES

Quelque opinion que l'on puisse avoir sur le caractère objectif des probabilités, force est de reconnaitre que nous rencontrons celles-ci d'abord dans la pensée du probabiliste, et le chapitre précédent nous a préparés à faire une grande place à l'étude de cet homme qu'est le statisticien vivant ; c'est donc dans cette pensée du statisticien que nous demandons la permission de commencer notre analyse, sans rien préjuger des conséquences auxquelles cette analyse pourra nous conduire. Il sera temps de tirer ces conséquences après que nous aurons dégagé les faits tels qu'ils se présentent d'abord.

Mais il nous faut auparavant attirer l'attention sur certaines imperfections de notre langage, imperfections dont nous sommes trop souvent dupes : les considérations qui suivent pourront paraitre relever de la métaphysique, mais elles sont en réalité simple affaire de vocabulaire.

Strictement parlant, il n'existe jamais deux réalités identiques : pareille expression est proprement absurbe, car si ces réalités étaient identiques, elles se seraient pas deux, mais une seule ; lorsque nous parlons d'identité entre un être A et un être B, c'est habituellement que nous ne parlons ni de A ni de B, mais d'un troisième être C, qui n'étant ni A ni B possède avec A et B certaines relations si intimes, que nous confondons dans notre langage C avec A et aussi avec B. Puissions-nous ne commettre pareille confusion que dans notre langage...

Si par exemple je pense durant un certain temps (mettons : deux secondes) au triangle rectangle, je dis que ma pensée est la même dans la première et dans la deuxième seconde : en rigueur il faudrait dire que dans ces deux secondes mes pensées ont avec l'idée de triangle rectangle des relations semblables (cette similitude étant elle-même une certaine relation d'espèce particulière et assez mystérieuse.)

De même si je soupèse durant deux secondes un objet de 1.000 grammes, mes sensations dans la première et dans la deuxième seconde ont entre elles une relation telle que je les déclare "équivalentes" : mais il serait peu correct et seulement approximatif de dire qu'elles sont identiques, qu'elles sont la même sensation "prolongée" (1).

---

(1) Si j'ai pris deux secondes, c'est naturellement pour la commodité du langage. Le problème se poserait d'une quantification du temps, au moins du temps psychologique mais l'unité élémentaire serait en général inférieure à la seconde. Ceci, toutefois, est une autre histoire ...

Outre les imprécisions du langage que nous risquons d'ériger en réalités objectives, il faut se garder d'un péril plus grand encore ; celui de projeter dans la conscience les réalités physiques ou physiologiques qui sont la cause de la sensation, mais qui ne sont pas la sensation ; nous sommes facilement abusés en cette matière par des schémas imaginatifs dont nous ne savons pas nous affranchir : nous dirons par exemple que notre sensation est l'intérieur, le dedans d'un état du système nerveux, façon de parler poétique qui serait légitime si le mot "dedans" n'évoquait à notre esprit l'image de nerfs creux comme des tuyaux. Ma sensation est sans doute conditionnée par un certain état de mon système nerveux lui-même causé par la force de traction du corps pesant ; mais il n'est pas permis a priori de passer sans précaution de la cause à l'effet, de la condition de la sensation à la sensation elle-même, pas plus qu'il n'est correct, en toute rigueur, de dire que durant deux secondes l'objet soupesé est le même ; si Zénon d'Elée avait pris cette précaution(toute verbale), il aurait vu que la flèche qui vole de A en B n'est pas un être, mais plusieurs êtres entre lesquels existe une certaine relation très intime (et encore aujourd'hui mystérieuse (1) ; c'est pour exprimer l'extrême intimité de cette relation que les hommes n'ont pas trouvé d'expression meilleure que de parler de "la même" flèche, ne sachant plus dès lors ce qui est désigné par le mot flèche ; est-ce la réalité partant de A ? celle qui arrive en B ? celle qui passe en quelque point N intermédiaire ? est-ce encore cette autre réalité qu'est l'histoire du mouvement qui s'est développé de A en B ? Notre langage mêle tout cela sous l'unique vocable "la flèche" ; ne tombons pas comme Zénon dans cette difficulté de langage ; et pour cela efforçons-nous de penser la réalité, au lieu de penser (tel Marius !) les mots qui désignent cette réalité, et la désignent plus ou moins heureusement...

Notre exemple des sensations produites par un corps pesant va nous introduire au sujet du présent chapitre.

Si à un objet de 1.000 grammes est ajouté au bout d'une seconde un poids de 2 grammes par exemple, nous ne nous en apercevons pas : dirons-nous que nous éprouvons deux sensations dont la différence nous échappe ? Voilà qui serait un exemple de cette fâcheuse erreur consistant à projeter dans notre sensation ce que nous savons de ses causes physiques.

En réalité il faut dire que les deux sensations (car elles sont bien deux : non parce que le poids a augmenté, mais parce que le temps s'est écoulé, ainsi que nous l'avons dit précédemment (2) les deux sensations sont équivalentes ; naturellement ne les disons pas identiques puisqu'elles sont deux. Mais elles sont équivalentes aussi bien et au même titre dans l'expérience où je soupèse durant deux secondes un poids de 1.000 grammes, et dans celle où l'objet soupesé passe de 1.000 à 1.002 grammes : la cause physique de cette équivalence n'est pas la même dans l'un et l'autre cas, mais l'effet psychologique est bien le même : ce qui n'est pas ressenti n'est aucunement sensation, c'est trop clair...

(1) Cependant il semble que la physique quantique nous met sur la voie, et l'on s'apercevra bientôt sans doute que décidément ce sont les "états" de Dirac qui ont une réalité objective (d'un aspect très différent de celui que nous prêtons actuellement à la réalité), tandis que les α , les "observables" sont une relation entre les ψ et l'observateur -relation que l'observateur a la fâcheuse habitude de "projeter" dans la réalité, d'objectiver ; non certes que les α n'existent pas, mais ils n'existent que là où se trouve un observateur qui produit les α par sa relation aux ψ . Ainsi se trouvera pleinement sauvegardé le déterminisme de la matière ; c'est notre liberté qui introduit l'indétermination dans notre connaissance de cette matière. Par ailleurs cette matière ne ressemble pas du tout aux images que nous nous en forgeons, et qui sont seulement des images de notre relation à la matière.

(2) Rappelons que c'est pour simplifier le langage que nous prenons la seconde comme "quantum" de temps psychologique.

Ainsi apparaît entre deux sensations une certaine sorte de relation que nous appelons "équivalence", et dont nous allons étudier les propriétés.

Il résulte des expériences classiques de Fechner que, étant donnés trois poids successifs α, β, γ il arrive que l'on puisse distinguer α de γ sans pouvoir cependant distinguer α de β ni β de γ

Il existe entre les poids physiques un "seuil" au-dessous duquel l'organisme d'un sujet donné ne sait pas faire de distinction. Evitant, comme il a été dit, de projeter les poids physiques α, β, γ dans les sensations A, B, C, nous devons dire qu'en ce qui concerne ces sensations, il existe ;

entre A et B, entre B et C, une relation d'équivalence,

entre A et C, une relation de distinction,

Il apparait ainsi que la relation d'équivalence entre réalités psychologiques n'est pas nécessairement transitive.

Que l'on ne se récrie pas au nom du principe d'identité ; nous avons suffisamment expliqué que la notion d'identité n'a rien à voir ici ; si on l'invoque c'est qu'encore une fois on projette dans le domaine psychologique un schéma qu'on s'est fabriqué avant d'observer la réalité psychologique telle qu'elle est. Par exemple, on a admis d'avance que les sensations sont mesurables par des nombres, les sensations équivalentes étant mesurées par des nombres égaux ; et comme l'égalité arithmétique est transitive on en conclut que l'équivalence psychologique doit l'être aussi : mais c'est précisément là toute la question.

En toute rigueur, voici la seule conséquence légitime que l'on puisse tirer du caractère transitif de l'égalité arithmétique : comme cette transitivité est regardée comme essentielle à la science mathématique, on ne peut établir entre l'univers et un schéma mathématique de correspondance biunivoque où l'équivalence soit représentée par une égalité, que si dans l'univers toute équivalence est transitive. Si l'univers contient en fait des équivalences non transitives, sa représentation par un schéma mathématique du type précédent est impossible. Autrement dit encore, un tel univers n'est pas totalement réductible à un schéma quantitatif. La loi expérimentale de Fechner semble bien démontrer que tel est le cas pour l'univers réel, car les faits psychologiques font partie de cet univers, et agissent sur lui (à une échelle que la bombe atomique a rendue appréciable).

Si l'univers contient des réalités "qualitatives" non réductibles à la quantité (1), cela ne signifie pas qu'on doive renoncer à en faire un étude scientifique et même parfaitement positive..On pourra même, si l'on veut, utiliser pour cette étude une espèce de mathématique de type nouveau, une "mathématique de la qualité" ; en commençant par dilater notre système de symboles "mathématiques" aux dimensions de l'univers, on pourra bien dire que l'univers est réductible aux mathématiques : ce n'est plus alors qu'affaire de langage.

Nous allons esquisser une telle algèbre de la qualité, prélude nécessaire, à notre avis, de la statistique mathématique.

ENSEMBLE « COMPARABLE »

Considérons un ensemble d'"êtres" A, A', C.. auxquels on a associé une certaine opération, nommée équivalence, qui se notera : $A \asymp A'$ (et qui est supposée identique à sa réciproque $A' \asymp A$).

Si l'équivalence est transitive, on peut faire correspondre à l'ensemble précédent un ensemble dit "isomorphe" X_1, X_2,... tel qu'à tout A corresponde un X, qu'à deux A non équivalents correspondent deux X distincts et qu'à deux A équivalents corresponde un seul X (1).

Mais si l'équivalence n'est pas transitive, cette réduction isomorphique n'est pas possible ; c'est parce qu'on a l'habitude de là postuler qu'on se scandalise à la pensée d'une équivalence non transitive.

L'équivalence peut d'ailleurs être transitive pour une partie seulement des A : dans l'avenir nous supposerons d'ordinaire que pour ceux-là la réduction a été effectuée. Supposons maintenant qu'en outre de l'équivalence soit définie une certaine relation nommée prévalence ($A \succ B$), possédant les propriétés suivantes : elle est exclusive de l'équivalence, ainsi que de sa réciproque (sa converse se nommera postvalence, notée : $B \prec A$) Bien que ce ne soit probablement pas toujours le cas, nous supposerons la prévalence transitive :

de $A \succ B$ et $B \succ C$ on tirera : $A \succ C$.

En général la prévalence ne commute pas avec l'équivalence, autrement dit de : $A \asymp B$ et $B \succ C$ on ne peut tirer $A \succ C$, car se serait exclure la possibilité d'avoir simultanément ; $B \asymp A$, $A \asymp C$, $B \succ C$: en ce cas l'équivalence serait transitive, ce qu'on ne suppose pas.

Si maintenant, étant donnés deux êtres quelconques d'un certain ensemble E, il existe toujours entre eux soit équivalence soit prévalence (ou postvalence), l'ensemble sera dit comparable. Eu égard à ces relations d'équivalence et de prévalence, les éléments de l'ensemble seront alors nommés des "grandeurs".

Ce nom de grandeurs ne doit pas nous faire illusion : il n'appartient pas aux éléments eux-mêmes, mais à l'association de ces éléments avec leurs relations de comparaison ; en outre, si l'équivalence n'est pas transitive dans l'ensemble (2) nous savons qu'on ne peut associer à chaque élément un symbole qui constituerait "sa grandeur", de telle sorte qu'à deux éléments équivalents corresponde "une même grandeur". C'est donc avec beaucoup d'hésitation que nous acceptons d'employer ce nom dangereux de "grandeurs" : il s'agit de grandeurs improprement dites. Si au contraire l'équivalence est transitive (2) le terme de grandeur ne fait plus de difficulté : nous parlerons alors de grandeurs vraies, et l'ensemble E sera dit ordonné.

Examinons un cas particulier important, celui d'un ensemble E comparable mais non ordonné (c'est-à-dire dans lequel l'équivalence n'est pas transitive) satisfaisant aux deux conditions suivantes :

1°) De $A' \asymp A$ et de $A \succ B \succ C$ on peut tirer $A' \succ C$.

2°) Etant donné un couple quelconque d'éléments équivalents $A' \asymp A$, on peut trouver des X équivalents à A' et prévalents à A, ou bien on en peut trouver d'équivalents à A et prévalents à A', mais non les deux à la fois.

---

(1) En ce cas nous emploierons les notations classiques : $=$, $<$, $>$ au lieu de $\asymp$, $\prec$, $\succ$.

(2) On notera qu'elle peut l'être dans un sous-ensemble de E sans l'être dans E ; auquel cas la réduction isomorphique effectuée dans le sous-ensemble n'est pas correcte dans l'ensemble total : source d'erreurs fréquentes.

Dans ces conditions, on peut ordonner A et A' par référence aux X, et cet ordre est transitif. L'ensemble E, dans ce cas, bien que non ordonné par nature, est ainsi ordonnable par un artifice. A chaque élément A on peut associer l'"intervalle" (ordonné comme il vient d'être dit) de ses équivalents. Si l'on a défini dans l'ensemble E une addition (comme nous l'expliquons ci-dessous) un tel ensemble est alors approximativement mesurable, c'est-à-dire qu'il est représentable par des points d'un axe avec un seuil E à l'intérieur duquel se trouvent les points figuratifs de l'intervalle associé à A (1). C'est ce qui se passe pour la loi de Fechner par exemple.

ENSEMBLE MESURABLE

Notre dessein n'est pas de faire ici une axiomatique en règle, mais seulement d'ouvrir la voie à un pareil travail. Aussi nous contenterons-nous d'esquisser quelques notions, d'énoncer quelques postulats sans chercher à être exhaustif. Il nous faut expliquer ce qu'est l'addition, qui nous conduira à la notion de mesure : bornons nous à quelques principes fondamentaux.

L'addition telle qu'on l'envisage dans le domaine des mathématiques classiques, est une opération qui associe à un couple d'éléments A, B, appartenant à un certain ensemble E un troisième élément C, cette opération étant astreinte à posséder certaines propriétés : il est habituellement demandé à l'addition d'être commutative et associative, d'admettre une opération inverse (soustraction) et un élément unitaire (le zéro) ; dire qu'il y a un zéro, c'est dire que l'ensemble E' des C contient l'ensemble E, ou même s'y réduit : s'il ne s'y réduit pas, on supposera qu'on peut encore définir d'addition dans l'ensemble E', puis dans l'ensemble E" qui en résultera, etc.. et l'on aboutira à un ensemble $\mathcal{E}$ (peut être infini ou transfini) dans lequel l'addition formera groupe : nous dirons que $\mathcal{E}$ forme un ensemble additif. D'autre part on a supposé dès le début que l'addition commute avec l'égalité, c'est-à-dire que, de : A' = A et : B' = B, on tire : A' + B' = A + B

Quand nous abordons les "grandeurs psychologiques", la commutativité et l'associativité de l'addition ne font généralement pas difficulté ; il n'en va pas de même des autres conditions.

Et d'abord, si l'équivalence (qui remplace ici l'égalité) n'est pas transitive, il est difficile de supposer que l'addition commute avec elle. On pourra bien parfois, de $A' \asymp A$, conclure $A' + B \asymp A + B$; mais la réciproque sera inexacte, ce qui rendra difficile une définition de la soustraction. Ces difficultés pourront être surmontées dans le cas des ensembles qui, sans être ordonnés, sont ordonnables indirectement comme il a été expliqué plus haut : il suffira que l'addition commute avec cet ordre.

Reste une dernière difficulté, sur laquelle nous attirons très particulièrement l'attention, car elle est souvent négligée : lorsqu'on part d'un ensemble E possédant certaines propriétés, l'ensemble E' des A + B peut avoir une structure fort différente, s'il ne se réduit pas à E lui-même. Et d'abord, il peut ne pas même contenir E : on se rendra compte en effet que souvent, ayant cru définir une "addition" des éléments de E, on a omis de s'assurer de l'existence d'un élément zéro. Ensuite, même si E' contient

(1) Nous n'avons pas examiné de très près si les conditions ici énoncées sont absolument suffisantes ; nous laissons ce soin à ceux que la question intéresserait.

E, l'ensemble $\mathcal{E}$ des E, E', E'', etc... peut être assez difficile à déterminer. Enfin on devra prendre garde que, partant d'un ensemble E comparable ou même ordonné, on peut aboutir à un ensemble E' qui ne sera ni ordonné, ni peut être même comparable.

Pour toutes ces raisons, nous ne définirons en principe l'addition que pour un ensemble $\mathcal{E}$ tout entier ordonné, ou au moins ordonnable : nous rentrons alors dans le domaine des mathématiques classiques. Moyennant quelques conditions supplémentaires bien connues (existence d'autant d'éléments que l'on veut équivalents à un élément donné A, possibilité de "dépasser" tout élément B par addition successive d'éléments équivalents à A, ou encore existence des "sous-multiples" pour tout élément) on pourra définir une mesure de $\mathcal{E}$, c'est-à-dire réaliser une isomorphie entre $\mathcal{E}$ et un ensemble de nombre réels, avec correspondance des opérations. Nous dirons alors que l'ensemble $\mathcal{E}$ est mesurable. Si l'équivalence n'est pas transitive, mais que nous ayons pu ordonner E par le procédé indirect exposé plus haut, nous dirons que $\mathcal{E}$ est approximativement mesurable ; dans ce dernier cas, on notera qu'il n'y a pas correspondance entre l'équivalence dans $\mathcal{E}$ et l'égalité des mesures, c'est l'inégalité $|x-x'| < \varepsilon$ qui est l'image de l'équivalence. (1)

NOTIONS DIVERSES

Nous dirons que E' est un "sur-ensemble" de E, lorsqu'il contient E, lors donc que E est un sous-ensemble de E'. Nous avons attiré l'attention sur la prudence avec laquelle on doit passer d'un ensemble à un sur-ensemble. Etudions le cas où, étant donnés deux ensembles mesurables E_1 (formé d'éléments A) et E_2 (formé d'éléments B) on associe à chaque couple (A, B) un élément C d'un ensemble $\mathcal{E}$.

Il peut se faire que les trois ensemble E_1 , E_2 et $\mathcal{E}$ coincident, et que l'association (A, B) ne soit pas autre chose que l'addition A + B. Il peut se faire que, E_1 , E_2 et $\mathcal{E}$ coincidant toujours, (A, B) constitue une nouvelle opération : on connait bien, en particulier, les conditions pour que cette opération soit dite "multiplication".

Mais il peut se faire encore que, E_1 , et E_2 étant ou non identiques, $\mathcal{E}$ soit un ensemble nouveau. Pour considérer l'association (A, B) comme une opération, nous exigerons qu'elle commute avec l'équivalence (supposée transitive, puisque E_1 et E_2 sont mesurables); cela soulèvera de gros problèmes en pratique, problèmes que l'on a trop souvent tendance à négliger comme d'avance résolus.

Que supposons-nous maintenant sur $\mathcal{E}$? Il n'est pas toujours nécessaire de le supposer mesurable, ni ordonné, ni même peut-être comparable. Parfois, nous saurons seulement que les sous-ensembles de $\mathcal{E}$ obtenus en fixant soit A, soit B, sont mesurables ; en ce cas, à certaines conditions on pourra regarder l'association (A, B) comme une sorte, soit d'addition, soit de multiplication, complexe.

Ces considérations auront suffi à montrer l'extrême complication de ces problèmes, et la prudence avec laquelle on doit avancer dans ce domaine, singulièrement plus délicat que celui des grandeurs algébriques.

---

(1) Nous ajournons la recherche des conditions complètes pour qu'un ensemble $\mathcal{E}$ soit approximativement mesurable ; question connexe : l'intervalle est-il à prendre ouvert ou fermé ?

Pour terminer, quelques mots sur une notation dont nous ferons grand usage :

Lorsqu'un ensemble E de grandeurs A_i est mesurable (en considération d'un système Σ d'opérations), pour exprimer que les A_i, sont entre eux comme les nombres α_i nous écrirons :

$$A_i \overset{\Sigma}{/E/} \alpha_i$$

ou plus simplement :

$$A_i / i / \alpha_i \quad \text{ou même} \quad A_i // \alpha_i$$

lorsqu'il n'y aura pas ambiguïté.

(Le symbole // ou /i/ pourra se lire :

"....proportionnel à .." ou "... proportionnel (i) à ...")

Si les A dépendent de deux indices, nous écrirons de même :

$$A_{ij} /ij/ \alpha_{ij} \quad \text{ou seulement :} \quad A_{ij} // \alpha_{ij}$$

Si les sous-ensembles à j constant sont seuls mesurés par certains nombres, et <u>si l'élément mesuré par zéro ne dépend pas de j</u>, nous écrirons:

$$Aij / i / \alpha_{ij} \quad \text{ou :} \quad Aij / i / \alpha_i$$

selon que les rapports des Aij à j constant dépendent ou non de cette valeur constante de j.

Lorsqu'il sera nécessaire de préciser une restriction (comme par exemple la constance de j) nous l'écrirons à la suite, séparée par un triple trait vertical :

$$Aij / i / \alpha_{ij} /// j = C^{te}; \text{ ou simplement : } /// j$$

Si les sous ensembles à i constant et à j constant sont mesurables(1) sans que peut-être l'ensemble total le soit, nous écrirons :

$$A_{ij} / i / j / \alpha_{ij}$$

On ne devra pas sans justification en tirer :

$$Aij / ij / \alpha_{ij}$$

comme on le fait trop souvent : <u>cela n'est correct que si l'on a établi que l'ensemble (A_{ij}) est mesurable</u>.

Eclairons les considérations qui précèdent par un exemple, qui a du reste pour nous une grande importance : la notion de <u>prix</u>, ou "valeur".

<u>Exemple : la notion de "valeur"</u>

On a coutume d'attacher les valeurs à des "choses" (objets matériels, énergie, situations..) ; cette façon de voir soulèverait des difficultés parfois inextricables, et ne permettrait pas souvent de mesurer les valeurs :

(1) Les éléments zéro restant toujours fixes.

les intérêts respectifs portés par un homme à un pain, une séance de cinéma, sa profession, sa femme..., ces intérêts ne paraissent en aucune manière comparables. En revanche si cet homme se trouve dans une situation déterminée ayant à prendre une décision pour l'instant qui vient, il peut avoir à choisir entre : manger maintenant, aller maintenant au cinéma finir maintenant un travail commencé, retourner maintenant au domicile où sa femme l'attend : "peser" ces décisions, cela a un sens ; c'est une psychologie de la décision que constitue donc une étude de la notion de valeur.

Dans ce chapitre, nous nous bornerons aux décisions dont les conséquences sont regardées comme certaines (et c'est pourquoi on a coutume de reporter les valeurs sur ces conséquences) mais nous étudierons plus loin des décisions dont les conséquences sont aléatoires, et force nous est en vue de ce cas, d'employer un langage plus rigoureux.

Soient d'abord deux décisions G et H qui s'excluent l'une l'autre (du moins dans l'immédiat où seul je les envisage ici), et entre lesquelles j'hésite. Lorsque je préfère G à H, j'écrirai : $G \succ H$, et $G \prec H$ dans le cas inverse. Si je n'ai pas de préférence, je dirai que les deux décisions sont équivalentes, et j'écrirai :

$$G \asymp H^{(1)}$$

Les trois relations précédentes ($\succ$ $\prec$ et $\asymp$) possèdent presque toujours les propriétés que nous avons énoncées comme caractéristiques des grandeurs comparables (2). Voici maintenant comment on peut définir l'addition, mais (comme nous l'avons dit) cette définition ne semble intéressante que si l'équivalence est transitive, ou si du moins on peut introduire un ordre indirect qui la rende "approximativement transitive" (ce qui est souvent le cas).

Soient G et H deux décisions compatibles. A certaines conditions psychologiques elles seront dites "indépendantes", et nous admettrons que la conséquence en est que la double décision G + H a une "valeur" qui ne dépend que des valeurs de G et de H, en d'autres termes : dire que des décisions sont indépendantes, ou dire que leur addition commute avec l'équivalence, ce sera à nos yeux deux énoncés (l'un psychologique, l'autre logique) d'une même réalité. Il importe de remarquer, dans ces conditions, que des décisions indépendantes dans un ensemble E peuvent cesser de l'être dans un sur-ensemble E' : cette conséquence nous parait très admissible, et très éclairante pour certains problèmes de psychologie.

Lorsqu'alors sont remplies les conditions convenables, et qu'en particulier les G, H..., les G + H, appartiennent à un même ensemble comparable, ordonné ou du moins ordonnable, on peut mesurer les valeurs des décisions par des nombres (de façon exacte ou seulement approchée) : nombres relatifs naturellement, définis à un facteur près, et que nous noterons ϖ (G) :

$$G \overset{\varpi}{/E/} \varpi(G) \qquad \text{ou plus simplement} \quad G//\varpi(G)$$

le ϖ placé au-dessus exprimant le point de vue d'où sont comparables les G

(1) Il est supposé que G et H sont envisagés sous un certain rapport, et qu'en écrivant $G \asymp H$ je me réfère à cette façon déterminée de les comparer (ainsi qu'à la personne qui fait cette comparaison).

(2) En fait la prévalence n'est sans doute pas toujours transitive : mais nous négligerons cette circonstance.

Il arrive souvent qu'un ensemble de décisions G_i étant mesurable selon le procédé qui vient d'être expliqué, et de même un ensemble H_j, la réunion de ces deux ensembles G_i et H_j ne sera plus mesurable ni même comparable. En ce cas, il arrivera d'ordinaire que les $G_i + H_j$ sont mesurables lorsqu'on fixe soit i, soit j. Cependant les éléments zéro n'étant pas fixes, on ne peut user de la notation exposée plus haut, mais sous certaines conditions on pourra représenter la valeur de $G_i + H_j$ par un vecteur $I \varpi (G) + J \varpi (H)$.

CHAPITRE III

VRAISEMBLANCES - PROBABILITÉS - ÉPREUVES

La notion de "valeur" nous a fourni un exemple de grandeur psychologique mettant bien en évidence les difficultés signalées au chapitre II. Abordons à présent la notion qui constitue l'objet propre de notre étude, la notion de "vraisemblance".

Indiquons d'abord que nous ne prenons pas ce mot exactement au même sens que R.A. Fisher, mais que nous désignons par là une grandeur analogue à ces étranges "probabilités a priori" de la formule de Bayes. Plus précisément nous montrerons que si l'on admet la notion que nous appelons "vraisemblance", on peut légitimer (dans certains cas) l'usage d'une formule semblable à la formule de Bayes, mais dans laquelle les premiers nombres mesurent des "vraisemblances" et non pas des probabilités. Nous verrons d'autre part que la notion de vraisemblance est antérieure à celle de probabilité, qu'elle la déborde, et qu'elle en rend compte. Si bien qu'en fin de compte cette notion, fort mystérieuse j'en conviens, mais absolument indispensable, constitue le véritable fondement et du calcul des probabilités, et de la théorie correcte de l'estimation statistique.

Fort mystérieuse, ai-je dit : on ne s'en étonnera pas, si l'on réfléchit que les notions premières de toute science présentent toujours ce caractère. Or nous pensons que le calcul des probabilités est une science originale, et non une simple branche des mathématiques. Et c'est faute d'avoir vu ou accepté cette originalité que l'on a vainement tenté jusqu'à ce jour (telle est du moins notre opinion) une axiomatique du calcul des probabilités ; on a voulu que ce dernier soit aux mathématiques ce que la géométrie est à l'analyse : on sait comment la géométrie analytique permet d'identifier formellement les être géométriques à des êtres analytiques. Mais à notre avis cela n'est pas totalement réalisable en calcul des probabilités. On peut bien réduire à la théorie de la mesure des ensembles les grandeurs qui mesurent les probabilités ; on ne peut pas y réduire la façon d'utiliser ces grandeurs. Dès que l'on veut qu'une probabilité signifie quelque chose, on fait éclater le cadre d'une axiomatique exclusivement abstraite : on le voit bien par exemple dans les déconcertantes questions de probabilités arithmétiques (sur lesquelles Monsieur Borel a si heureusement attiré l'attention).

A ceux que heurteraient les idées que nous allons exposer maintenant, nous demanderons de nous faire provisoirement crédit, comme d'Alembert le demandait pour l'usage des imaginaires (dont il fallut attendre un siècle la pleine justification logique) : "employez-les toujours, la foi vous viendra..."

VRAISEMBLANCES ET PROBABILITÉS

Lorsque je (1) me trouve en présence d'un problème concret à résoudre : "Y a-t-il du feu dans cette maison là-bas ?" etc..., l'ensemble de mes connaissances passées et de mes informations présentes (peut être aussi la structure particulière de mon cerveau et mes dispositions sentimentales) tout cela concourt à me suggérer diverses réponses, que je juge diversement "vraisemblables" ; Il y a de la fumée, il doit y avoir du feu. -Ce que je prends pour de la fumée n'est peut être que du brouillard.- Il est midi, la maison parait habitée, il est donc vraisemblable qu'on y prépare le déjeuner -Ou même : je suis trempé, ma bonne étoile a dû me favoriser, une fois de plus et me préparer du feu -ou au contraire : Avec ma malchance habituelle je ne trouverai pas à me sécher - Etc...

Ainsi dans une circonstance bien déterminée, à diverses hypothèses touchant la réalité, j'attribue des degrés divers de "vraisemblance".

Ces "vraisemblances" sont des grandeurs psychologiques ; elles sont relatives à un problème objectif jugé par une sujet. On ne peut donc les qualifier a priori ni de purement objectives, ni de purement subjectives : elles se présentent comme essentiellement relatives.

Est-ce que a postériori on pourrait montrer qu'elles ne dépendent du sujet qu'en apparence, et que tout sujet ayant les mêmes informations les appréciera de la même façon ? Il ne nous semble pas : les connaissances passées du sujet y jouent un rôle irréductible. Dira-t-on alors qu'elles seront du moins appréciées semblablement par deux sujets ayant strictement les mêmes connaissances passées jointes aux mêmes informations présentes ? Nous ferons remarquer que ces deux sujets ne seraient pas alors deux, mais un seul : cependant nous admettrons volontiers que pour un tel sujet, il n'y a effectivement qu'une seule façon légitime d'apprécier les vraisemblances, et que s'il s'en écarte c'est qu'il fait une place (indue) à ses dispositions sentimentales. En ce sens-là, on peut parler du caractère objectif des vraisemblances : nous dirons donc qu'elles sont objectivement relatives, qu'elles sont une relation objective d'un sujet à un objet. En outre il arrivera souvent que tous les hommes placés dans une même situation apprécieront les vraisemblances de façon pratiquement identique. Mais nous ne croyons pas devoir retenir la notion d'une vraisemblance qui serait objective en ce sens qu'elle serait attachée à l'objet seul : cette remarque (qui nous parait de simple bon sens) est très importante, car elle condamne tout une façon contemporaine d'interpréter la mécanique quantique.

Soient maintenant A, B, C,.. diverses hypothèses que j'envisage concernant un certain problème concret Z - ce problème étant : Y a-t-il du feu dans cette maison ? Ou encore : la pièce de monnaie que j'aperçois de loin sur cette table est-elle tournée côté face, ou côté pile ?

A chaque hypothèse j'attache, comme il vient d'être dit, une "vraisemblance", grandeur psychologique que nous noterons V(A), V(B)...(2) On notera à ce sujet que :

---

(1) Dans tout ce qui suit, le statisticien est un homme bien déterminé. Pour alléger le style, il sera commode de le faire parler à la première personne.

(2) S'il est nécessaire d'expliciter le fait que cette vraisemblance est relative à un sujet déterminé S (nous entendons par là ce sujet pris avec tout ce qui le spécifie : connaissances passées, informations présentes, problème Z qu'il cherche à résoudre) nous noterons : V (AS). Il est en effet d'un haut intérêt de remarquer que l'opinion d'un sujet S sur une proposition hypothétique A possède certaines propriétés de la multiplication.

1° - Il n'est pas nécessaire pour donner un sens à V (A) d'envisager toutes les hypothèses possibles ; V(A) n'est pas la proportion de vraisemblance que j'attribue à A parmi ces diverses hypothèses, mais la vraisemblance que j'attribue directement à la seule hypothèse A,

car 2° - V (A) n'est pas un nombre, mais une de ces mystérieuses "grandeurs psychologiques" dont nous avons déjà parlé.

En fait, je ne suis peut être pas même absolument certain qu'il y ait une pièce de monnaie sur la table là-bas. S'il n'y en a pas, si ce que je prends pour une pièce n'est qu'une tache sur la nappe, la question serait dépourvue de sens. Elle ne l'est pourtant pas ; il suffit de formuler A ainsi: y a-t-il sur cette table une pièce de monnaie côté face ? B se formulera de même : y a-t-il sur cette table une pièce de monnaie côté pile ?

Ainsi V (A) et V (B) ont chacune un sens propre, indépendamment de leur comparaison. Encore une fois, le caractère mystérieux de cette notion de vraisemblance ne l'empêche aucunement d'exister, et n'empêche pas davantage d'en faire une certaine étude, qui pour être imparfaite n'en est pas moins importante et instructive.

Supposons maintenant qu'il soit possible de comparer V(A) et V(B), selon ce qui a été expliqué au chapitre II. En particulier, dans l'exemple précédent, si je suis assez loin de la pièce que je crois apercevoir, je serai conduit généralement à attribuer à A et à B des vraisemblances qu'il m'est possible de déclarer équivalentes ; ce qui se notera : $V(A) \asymp V(B)$.

Lorsque l'équivalence est transitive (ce qui est du reste toujours le cas s'il n'y a que deux hypothèses envisagées) on peut écrire :

$$V (A) = V (B), \text{ ou : } V (A), V (B) /\!/ 1.$$

On définit l'addition logique de deux hypothèses A et B, comme étant l'hypothèse C = A + B selon laquelle de A et de B l'une au moins est vraie. Sous certaines conditions, concernant notamment l'incompatibilité de A et B, il est légitime de définir l'addition des vraisemblances par :
$V (A) + V (B) = V (A + B)$ (1). Il est alors possible d'écrire :

$$V (A), V (B), V (A + B) /\!/ 1,1,2.$$

On fera donc tout naturellement la convention :

$$\frac{V (A)}{V (A+B)} = \frac{V (B)}{V (A+B)} = \frac{1}{2}$$

Le quotient $\frac{V (A)}{V (A+B)}$ lorsqu'il est ainsi mesurable par un nombre recevra le nom de probabilité.

On remarquera à ce sujet deux choses :

1° - La restriction : "lorsqu'il est mesurable par un nombre". La probabilité (qui n'est pas une vraisemblance, mais un quotient de deux vraisemblances) est une notion de portée restreinte ; on peut toujours parler de vraisemblance, on ne peut parler de probabilité qu'à certaines conditions. En pratique on pourra souvent admettre que les vraisemblances, sans être strictement mesurables, le sont de façon approchée : cela suffira pour définir, de façon également approchée, leurs quotients. Il arrivera alors qu'un schéma de probabilité strict n'existant pas, on se conten-

(1) Voir au chapitre précédent ce qui a été dit sur la définition de l'addition.

tera d'un schéma approximatif : il ne faudra pas oublier ce caractère d'approximation, qui expliquera bien des discordances entre "l'esprit de géométrie" et "l'esprit de finesse".

2° - Nous n'avons pas eu besoin de supposer que A + B couvre la totalité des hypothèses possibles : C'est par un libre décret de ma volonté que j'ai décidé de ne m'intéresser qu'à elles deux. Si par exemple j'ai fait un pari avec un camarade, il est sous-entendu que s'il n'y a pas réellement de pièce sur la table le pari est annulé.

Nous sommes dès lors conduits à la définition suivante de la probabilité :

La probabilité d'une hypothèse A par rapport à un ensemble convenu d'hypothèses H est le quotient de la vraisemblance de A par la vraisemblance de H, lorsqu'on peut légitimement mesurer ce quotient par un nombre.

Comme on peut généralement se ramener aux cas de vraisemblances deux à deux équivalentes, nous dirons encore :

La probabilité d'une hypothèse A est le rapport entre le nombre de sous-hypothèses incompatibles qui constituent A et le nombre de toutes celles que l'on a décidé de retenir, lorsque ces diverses hypothèses peuvent être regardées comme également vraisemblables.

Nous retrouvons ainsi la définition classique, avec les différences suivantes :

1° - Nous ne parlons pas de probabilité d'un événement, mais de celle d'une hypothèse.

2° - A la formule parlant de "tous les cas possibles", nous en substituons une qui parle seulement des hypothèses que l'on a décidé de retenir.

3° - Nous faisons une restriction sur la légitimité d'une comparaison quantitative entre les diverses vraisemblances : pour autant que cette comparaison quantitative sera légitime, nous disons que le système est mesurable.

4° - En substituant l'expression "également vraisemblables" à l'expression "également probables" de la définition traditionnelle, nous la libérons d'un cercle vicieux bien connu ; on ne peut définir probabilité par probable, mais on peut définir probabilité par vraisemblable. On dira sans doute que nous n'avons fait que déplacer la difficulté ? Certes : mais que peut-on faire d'autre en matière de définition, que de remonter vers des notions de plus en plus primitives ? Or, nous pensons que la notion de vraisemblance, beaucoup plus large que celle de probabilité, lui est antérieure : en ramenant l'une à l'autre, nous n'avons pas supprimé une difficulté qui ne saurait être supprimée ; nous l'avons seulement mieux localisée ; n'est-ce pas déjà quelque chose ?

*

* *

Nous avons parlé dans ce qui précède de l'addition logique de deux hypothèses, et indiqué que sous certaines conditions cette opération permet de définir l'addition des vraisemblances. De même le produit logique d'hypothèses A et B permet, sous certaines conditions encore, de fonder la multiplication des vraisemblances.

Lorsqu'en effet les A_i et les B_j sont indépendantes, c'est-à-dire lorsque la connaissance éventuelle d'un A ne modifie pas les vraisemblances relatives des B, et réciproquement, (1) les vraisemblances des hypothèses doubles $A_i\ B_j$ sont entre elles comme celles des B_j lorsqu'on fixe A_i, comme celles des A_i lorsqu'on fixe B_j. Ce que nous écrivons :

$$V\ (A_i\ B_j)/i/j/V(A_i) \wedge V\ (B_j).$$

Si alors les $V\ (A_i\ B_j)$ sont mesurables, on sait qu'il en résulte : $V\ (A_i\ B_j)/i\ j/\ V\ (A_i)\ V\ (B_j)$

Cette proportion ne permet pas, remarquons-le, de comparer : $V\ (A_i\ B_j)$ à $V\ (A_i)$ ou à $V\ (B_j)$ (2) mais cela n'est généralement pas nécessaire, il suffit de pouvoir comparer les $V\ (A_i,\ B_j)$ entre eux.

Si les A_i et les B_j ne sont pas indépendants, il n'existe pas de formule simple donnant la vraisemblance $V\ (A_i.\ B_j)$; pour retrouver la formule classique sur la probabilité $p\ (A_i.\ B_j)$ il faudrait ajouter certaines conditions, généralement omises parce qu'on les regarde comme allant de soi, ce qui est discutable. (3)

ÉPREUVES

Nous sommes maintenant en mesure d'étudier ce qu'est une épreuve, et en quel sens elle fournit de l'information.

Soit, pour fixer les idées, une urne que nous supposons contenir une infinité de billes blanches et rouges dans une proportion inconnue. $\lambda, 1-\lambda$.

Nous prélevons un échantillon X de n billes, dont k se trouvent être blanches, et donc n - k rouges. Que peut-on dire au sujet de λ ?

Nous commencerons par supposer que nous avons, avant tout échantillonnage, quelque idée relativement aux valeurs possibles de λ, cette idée se traduisant par une fonction de vraisemblance $V\ (\lambda)$ (que nous n'avons besoin de connaitre qu'à un facteur près : désormais, sauf mention expresse nous appelons $V\ (\lambda)$ des nombres proportionnels aux vraisemblances psychologiques supposées mesurables).

Plaçons-nous avant que la composition de l'échantillon X ne soit connue ; il est absolument indifférent, du reste, que cet échantillon ait été déjà tiré ou non ; ce sont nos informations qui comptent, non pas la date à laquelle se situe l'évènement (sauf dans le cas où la date en question constituerait déjà un début d'information sur la question que je me pose). Je puis alors me poser une double question : composition de l'urne, composition $\frac{k}{n}$ de l'échantillon X. Cette double question donne lieu à une série d'hypothèses composées, que je puis caractériser par le couple de nombres (λ, k). A chacun de ces couples correspond, selon mes informations avant connaissance de X, une vraisemblance que je caractérise par son nombre proportionnel : $V\ (\lambda, k)$.

---

(1) Je ne suis pas certain que cette réciproque aille de soi.

(2) On a toutefois évidemment : $V\ (Ai\ Bj) \lesssim V\ (Ai)$

(3) Indiquons une notation commode : pour représenter la proposition qui énonce que B est vrai si A l'est déjà, écrire $\frac{B}{A}$. On a ainsi : $A\ B \equiv A\frac{B}{A}$, $V\ (AB) = V\ (A\frac{B}{A})$ Il est exact que $V\ (AB)/B/V\ (\frac{B}{A})$ IIIA, mais il est inexact en général que : $V\ (AB)/A/V\ (A)$ ///B . On remarquera que le syllogisme en "barbara" consiste à tirer de $\frac{B}{A}\ \frac{C}{B}$ la proposition $\frac{C}{A}$.

Au reste, certaines hypothèses que nous sommes décidés à admettre sur le caractère aléatoire des tirages et leur indépendance, nous permettent d'écrire :

$$v(\lambda, k) \;/\!/\; V(\lambda)\; C_n^k\; \lambda^k (1-\lambda)^{n-k}$$

Telle est la vraisemblance d'un couple (λ , k) lorsqu'on ne possède aucune information supplémentaire.

Plaçons-nous à présent après que nous avons pris connaissance de la composition de l'échantillon X. Le seul changement qui en résulte pour nos vraisemblances est que désormais $k = k_o$ est connu, et donc les seuls couples (λ, k) à considérer sont les (λ , k_o). Mais puisque nous admettons l'indépendance des vraisemblances de k et de λ , les vraisemblances relatives de ces couples sont les mêmes qu'avant notre information sur k, autrement dit :

$$v(\lambda, k_o) \;/\lambda/\; V(\lambda)\; C_n^{k_o}\; \lambda^{k_o} (1-\lambda)^{n-k_o}$$

ou simplement :

$$v(\lambda, k_o) \;/\!/\; V(\lambda)\; \lambda^{k_o} (1-\lambda)^{n-k_o}$$

Ainsi la connaissance de l'échantillon X a pour effet de multiplier les vraisemblances initiales V (λ) par le "facteur de vraisemblance" $\lambda^k(1-\lambda)^{n-k}$: ce facteur n'est pas à proprement parler une vraisemblance, mais il constitue un opérateur qui transforme les vraisemblances avant épreuve en vraisemblances après épreuves (1).

La même théorie s'applique immédiatement aux échantillonnages portant sur une loi continue. Soit f (x, m) la densité de probabilité d'une telle loi, où le paramètre inconnu m est regardé avant toute épreuve comme ayant une vraisemblance V (m)

Considérons un échantillon de n points x_1 , $x_2 \ldots x_n$; pour alléger l'écriture, nous le désignerons par le point figuratif dans l'espace à n dimensions, et conventionnellement f (ξ) sera mis pour $f(x_1) f(x_2) \ldots f(x_n)$.

Avant épreuve, nous pouvons considérer la vraisemblance de l'ensemble (m, ξ), laquelle est V (m) f (ξ) (étant entendu qu'à tout point ξ est associé un volume infiniment petit $d\omega$ qui ne jouera aucun rôle dans la suite, puisque les vraisemblances ici considérées sont toujours définies à un facteur commun arbitraire près).

Après n épreuves, les vraisemblances (à un facteur commun près) ne sont pas modifiées, si nous supposons m indépendant de ξ . Donc on a

$$v(m, \xi_o) \;/m/\; V(m, \xi)_{|||\,\xi=\xi_o\,|||} \;/\!/\; V(m)\, f(m, \xi_o)$$

Ce que Fisher appelait "vraisemblance" apparait donc ici encore comme un opérateur, ou "facteur de vraisemblance". Nous le nommerons souvent la "densité" correspondant à l'échantillon ξ ; cette densité devient égale à la vraisemblance nouvelle si l'on admet pour la vraisemblance initiale V (m) une valeur indépendante de m :

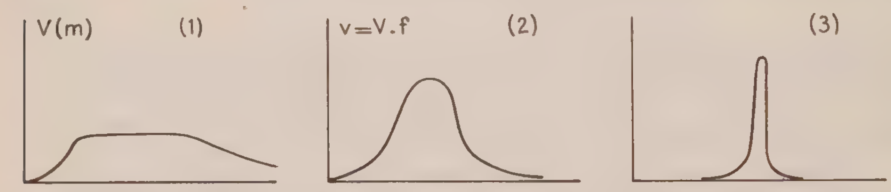
$$V(m) \;/m/\; 1$$

(1) En réalité, ce sont seulement les proportions de ces vraisemblances qui sont ainsi transformées : il serait peut-être malaisé de comparer une v avec une V (λ), mais nous n'en avons pas besoin.

(On remarquera que cette formule a un sens, même si l'intervalle où m est défini est illimité, tandis qu'en pareil cas on ne pourrait regarder V (m) comme une densité de probabilité ; c'est une des objections classiques contre la formule de Bayes ; on voit que dans notre conception cette objection ne porte pas).

On observera que dans ce qui précède nous n'avons pas une seule fois parlé d'estimation des paramètres λ ou m : c'est que le problème d'estimation, comme nous l'avons expliqué au chapitre I, n'a de sens strict que par rapport à un client c'est-à-dire par rapport à un système de valeurs, de prix attachés aux résultats ; nous aborderons ce problème au chapitre suivant. Pour le moment nous nous bornerons à tirer la conclusion de ce qui précède. Pour la simplicité du langage, nous nous référerons au cas d'une loi continue à un paramètre f (m, x).

Avant toute épreuve, nous admettons pour fonction de vraisemblance V (m), que nous pouvons représenter par une courbe ; par exemple celle de la figure (1). Après épreuve, " la fonction de vraisemblance devient comme nous l'avons vu, $V(m)\ f(m, \xi_o)$ et peut se représenter par exemple par la courbe 2. Si le nombre de points dans l'épreuve ξ_o est très considérable, la courbe représentative prendra l'aspect de la figure 3.



Lorsqu'ainsi on ne fait intervenir aucune considération de valeur du résultat, cela n'a pas grand sens de chercher dans l'estimation de m un nombre privilégié, sinon lorsque n devenant extrêmement grand, la courbe représentative se réduit pratiquement à une simple pointe. Ce cas excepté, il faut se résigner à faire correspondre à une épreuve ξ, non point un nombre, mais bien la courbe $v = V.f$ tout entière. A ce sujet, voici une terminologie qui nous semble commode :

Lorsqu'à un nombre ou un ensemble fini de nombres on fait correspondre un nombre qui en dépend, on dit que ce dernier constitue une fonction du ou des premiers :

$$x = f(y) \qquad x = f(y, z, t) \qquad \text{etc} \ldots$$

Lorsqu'à une courbe, à une fonction ou à un être quelconque mathématique (ou même physique) complexe on fait correspondre un nombre, on dit que celui-ci constitue une fonctionnelle, ou "nombre fonction de courbe".

Mais il n'existe pas de mot pour désigner inversement une courbe (ou tout être complexe) attachée à un nombre variable (ou éventuellement à un être plus compliqué, qui pourra lui-même être une fonction). Il est commode d'introduire un mot spécial pour ce cas très fréquent en statistique ; et comme un exemple classique est constitué par la fonction caractéristique, fonction attachée à une fonction, nous proposons d'employer le terme général de "caractérielle".

Ainsi la courbe $v = V(m)\ f(m, \xi_o)$ qui correspond à un échantillon ξ_o déterminé sera dite une caractérielle de ce même échantillon. Comme cette

caractérielle jouera un rôle décisif dans la théorie de l'estimation, nous l'appellerons la caractérielle estimatrice ou simplement : *estimatrice*. Pour une loi à deux paramètres, il va de soi que les estimatrices sont des surfaces, caractérielles à deux dimensions , etc...

Quand le nombre de points de l'échantillon est élevé, l'estimatrice devient très vite presque symétrique. Dans ces conditions son mode, sa moyenne, sa médiane, etc... sont pratiquement confondus, et comme nous verrons au chapitre suivant que ce sont des fonctionnelles de l'estimatrice de ce genre qui interviennent dans presque tous les problèmes d'estimation, presque toutes les méthodes d'estimation *fondées sur l'estimatrice* sont très vite pratiquement équivalentes. Ce n'est donc pas l'usage du *mode* de l'estimatrice qui est privilégié dans la méthode de Fisher, mais c'est l'utilisation d'une *fonctionnelle* de l'estimatrice ; au contraire, on aura des résultats moins bons en utilisant une fonction de l'échantillon ξ_o qui ne soit pas une fonctionnelle de l'estimatrice.

Par ailleurs on notera que V (m) étant d'ordinaire très aplatie, dès que n est un peu grand l'estimatrice est pratiquement indépendante de la vraisemblance initiale ; et l'on se contentera généralement de choisir celle-ci homogène. Alors la vraisemblance après épreuve se confondra avec la densité $f(m, \xi_o)$.

CHAPITRE IV

LA STATISTIQUE - SCIENCE DES DÉCISIONS

Nous avons noté comment, en tenant compte de certaines hypothèses qui l'ont amené à retenir un type de lois de probabilité f(m,x), de certaines vraisemblances initiales v(m), enfin de la connaissance d'un échantillon ξ_o le statisticien a établi, pour le paramètre m à estimer, une courbe estimatrice $v = \varphi(m) = v(m)f(m, \xi_o) = V(m) f(m, x_1) \ldots f(m, x_n)$.

Nous avons noté que lorsque n est élevé, l'estimatrice devenant d'abord symétrique, puis très pointue, toute estimation est bonne qui consiste à choisir m' pratiquement au centre de l'estimatrice.

Supposons maintenant que φ (m) étant quelconque, on se propose de faire pour m' le choix <u>le meilleur</u>. Quel sens a ce problème ?

En se reportant aux remarques de notre premier chapitre, on voit immédiatement qu'ainsi énoncé le problème n'a aucune signification, parce qu'il peut recevoir n'importe laquelle.

Toute la question est de savoir ce que désire le "client", et ce n'est donc pas aux statisticiens de décider du choix de l'optimum. C'est dire qu'il n'y aura pas <u>une</u> estimation optimum, mais autant qu'il y aura de types d'optimum désirables par le client.

Cependant, il faut noter que le statisticien a, dans cette question, un rôle à jouer, qui est d'éclairer son client et de l'aider à exprimer en langage statistique le désir qu'il ne saisit généralement lui-même qu'assez confusément. Or cela est difficile comme nous allons le voir.

A. - VALEURS ALÉATOIRES

Nous avons terminé le Chapitre II par une brève étude de la notion de prix, et nous avons souligné alors que les prix, mesures de valeurs lorsque les valeurs sont mesurables, étaient strictement parlant attachés aux décisions et non aux <u>objets</u> que ces décisions visent à obtenir. Cela nous a déjà permis de marquer le caractère <u>relatif</u> des prix, variables avec les individus, variables avec les temps et les circonstances pour un même individu. Cela va nous permettre à présent d'étudier le cas où une décision à prendre est de telle nature qu'elle n'obtienne pas infailliblement le résultat qu'elle poursuit et que l'on connaisse ce caractère "aléatoire".

Lorsqu'il en est ainsi, l'espérance que l'on a d'obtenir un certain résultat, la crainte aussi que l'on éprouve d'en obtenir peut-être un tout différent, tout cela met en jeu deux sortes de grandeurs psychologiques : la valeur du résultat, c'est-à-dire la valeur que nous attacherions à notre décision si nous jugions telle ou telle conséquence certaine : ce que nous nommerons sa valeur subjective - ou plutôt ses valeurs subjectives - si nous envisageons la possibilité de diverses conséquences et puis les vraisemblances de ces conséquences. C'est l'ensemble de tout cela qui contribue à former la valeur aléatoire de ma décision.

Cet ensemble est extrêmement complexe. Il n'est pas du tout certain qu'on puisse toujours dans l'étude d'une valeur aléatoire séparer vraisemblances et valeurs subjectives. Cependant, nous pensons que pour un homme capable de penser droitement, lucidement, sans se laisser dominer indûment par des considérations sentimentales, une telle séparation est possible : peut être même est-ce là un critère logique de degré d'objectivité d'un esprit ou de son "sentimentalisme" ; non que les "sentiments" n'aient pas de place chez un esprit objectif, mais cette place est celle des valeurs subjectives, elle ne doit pas déborder sur les vraisemblances, car un résultat n'est pas rendu plus ou moins vraisemblable par le désir que j'ai de l'obtenir ou de l'éviter (1).

Supposons maintenant que nous soyons en présence de vraisemblances mesurables et de valeurs subjectives également mesurables. Le problème que nous nous posons ici est de savoir si dans ces conditions les valeurs aléatoires résultantes sont,elles aussi, mesurables et, dans l'affirmative , que peut-on dire de leurs mesures ?

La réponse à la première partie de la question ne nous semble pas pouvoir être toujours affirmative. Un exemple simple est celui d'un billet de la Loterie Nationale : les valeurs subjectives des petits lots au moins sont manifestement mesurables (il n'en va probablement pas de même pour les gros lots ; ceux-ci restent du moins comparables) ; les vraisemblances sont très exactement mesurables aussi. Cependant, l'embarras de celui qui se demande s'il va risquer 300 francs certains pour un résultat aléatoire, l'embarras plus grand encore de celui qui se demande sur quelle série, A ou B, il va jouer : cela démontre assez que les valeurs aléatoires ne se mesurent pas exactement, surtout lorsqu'on songe que les espérances mathématiques (sur lesquelles nous reviendrons tout à l'heure) sont ici tout à fait désavantageuses.

On voit sur cet exemple le caractère absolument original de la valeur attachée à une décision à conséquences aléatoires.

Nous allons maintenant nous borner au cas où pour quelque raison, on peut considérer un ensemble de valeurs aléatoires comme mesurable. S'il en est ainsi (mais ce sera souvent difficile à établir) et si en outre (et ceci non plus ne va pas de soi) ces valeurs aléatoires sont complètement déterminées par la connaissance des valeurs subjectives et des vraisemblances, c'est-à-dire restent équivalentes à elles-mêmes lorsqu'on remplace les vraisemblances ou les valeurs subjectives par d'autres équivalentes ; alors il est facile de les calculer.

Traitons d'abord le cas où chacune des décisions étudiées D ne peut avoir qu'une seule conséquence intéressante pour moi, A, et celà avec une vraisemblance V. Il est clair que deux décisions de même conséquence

(1) Sauf naturellement le cas où je me demande quelle vraisemblance il y a que je prendrai une certaine décision.

agréable A (1) sont ordonnées comme les vraisemblances qu'elles ont de produire A. On peut même montrer que les valeurs de ces décisions sont additives comme leurs vraisemblances et finalement ces valeurs sont mesurables proportionnellement aux V :

$$\underset{\varpi}{} D_i / i / V_i \quad /// \ A$$

On voit de même qu'à vraisemblance constante, les décisions sont entre elles comme les prix ϖ_j des A

$$D_j / j / \varpi_j \quad /// \quad V = Cte$$

Ces deux proportions réunies s'écrivent (2) :

$$D_{ij} / i / j / \ V_i \wedge \varpi_j$$

Or, nous savons que si les D_{ij} forment un ensemble mesurable (et c'est ici notre hypothèse) on peut en tirer :

$$D_{ij} / ij / V_i \, \varpi_j$$

C'est la formule des "espérances mathématiques" : on voit les conditions très restritives qui la légitiment.

Supposons maintenant que la décision D puisse avoir deux conséquences A et B, l'une et l'autre m'intéressant (3). Sous certaines conditions (encore une fois nous renonçons ici à expliciter une axiomatique complète) on verra que si les valeurs subjectives de A et de B sont additives, il en ira de même des valeurs aléatoires. Ainsi sera complètement légitimé (avec des restrictions nombreuses) l'usage des espérances mathématiques, élargies du reste aux cas où l'on envisage des vraisemblances : une simple division ramènera au cas classique des probabilités.

B. - SIGNIFICATION DES VRAISEMBLANCES

Voici devant nous, sur une table un peu éloignée deux pièces de monnaie. Nous nous demandons comment elles sont tournées, ou plus exactement nous nous intéressons à la seule question suivante : combien y en a-t-il de tournées coté pile ? (quelles que soient celle ou celles ainsi tournées).

Dans les circonstances où nous nous trouvons, ce nombre x peut être 0, 1 ou 2, et nous attachons à ces trois nombres des vraisemblances formant un ensemble mesurable, et proportionnelles, dans cet ensemble, à 1, 2 et 1. Du reste, ce qui nous intéresse est simplement l'exactitude du jugement ; en sorte que la valeur subjective de chaque nombre possible pour x est la même. Nous sommes dans le cas où les valeurs aléatoires sont entre elles comme les vraisemblances, donc ici comme 1, 2 et 1. Il résulte

---

(1) L'ordre serait inversé pour une conséquence désagréable.

(2) A condition du moins que l'équivalence sur les D commute avec l'équivalence sur les V et sur les ϖ .

(3) On sera conduit à des prix positifs ou négatifs selon qu'il s'agit de conséquences désirées ou redoutées.

alors de la définition même de la valeur, que nous devons DECIDER d'admettre pour x la valeur 1, puisque c'est à pareille DECISION qu'est attachée la plus grande valeur.

Il nous faut très soigneusement refléchir sur ce résultat, qui renferme l'essence même du calcul des probabilités et de la statistique, et dont les répercussions sont considérables sur toute la théorie de la connaissance. La difficulté sera d'éviter deux interprétations inexactes : l'une par excès, l'autre par défaut, et de nous maintenir dans le juste milieu d'une pensée à la fois précise et nuancée : la tentation est si grande de sacrifier l'une à l'autre soit la précision, soit la nuance. Evitons de confondre la nuance avec la brume et la précision avec l'abstraction, ou finesse avec complication et simplicité avec simplisme. L'auteur demande toutefois quelque indulgence s'il n'a pas su réaliser ce difficile idéal...

L'adoption pour x de la valeur 1 résulte, avons-nous dit, d'une décision, Nous ne sommes aucunement assurés que x soit effectivement égal à l'unité et cependant nous devons faire comme si cela était vrai, du moins dans certains types de problèmes. Il ne s'agit donc pas ici d'un acte direct de connaissance, mais d'abord d'un acte de volonté : nous adoptons une hypothèse qui se présente avant tout comme préférable aux autres, d'une hypothèse que nous devons choisir en vertu d'une exigence de cohérence interne de notre esprit.

Il sera commode de donner un nom à ce processus intellectuel, où la volonté joue un rôle essentiel ; nous le nommerons une "assomption" ; la proposition "x égale 1" n'est pas saisie par connaissance directe, elle est "assumée".

Une proposition assumée peut n'être pas exacte ; la tentation est même grande de dire qu'elle n'a aucune raison de l'être et d'en venir à une conception strictement idéaliste, pragmatiste, de la statistique ; il n'y aurait alors aucune "connaissance" statistique, mais simple suppléance de la connaissance ; comme "il faut bien vivre", chaque fois qu'une connaissance proprement dite ferait défaut, on lui substituerait un schéma purement subjectif qui, impuissant à nous rien dire sur le monde, nous enseignerait ce que nous devons faire. Sur quoi se fonderait ce devoir ? Précisément sur l'exigence de cohérence interne de notre esprit qui, nous l'avons vu, règle le choix de telle ou telle assomption.

Nous pensons que, pour voir clair dans cette matière difficile, il faut distinguer deux questions : la première concerne le PROCESSUS selon lequel s'opère une assomption, la seconde concerne la PORTEE, la VALEUR de cette assomption.

Il est manifeste que, à prendre les choses comme nous les avons exposées (et nous pensons que c'est bien ainsi quelles se passent) le processus est, en effet, du type pragmatiste ; de ce point de vue, l'objet de la statistique mathématique et du calcul des probabilités est de fonder sur des principes cohérents les décisions que nous prenons, en fonction de certaines vraisemblances et des valeurs que nous attachons à l'obtention de certains résultats ; la statistique est bien la science qui assure à nos décisions une certaine cohésion logique. Plus particulièrement, c'est la science qui assure cette cohésion dans les cas où vraisemblances et valeurs constituent (au moins approximativement) des ensembles mesurables.

S'ensuit-il pour autant que ces assomptions soient dépourvues de toute valeur objective, quelles ne comportent aucune connaissance ?

De ce que l'assomption se fait selon un processus pragmatiste, faut-il conclure nécessairement que la portée réelle en soit nulle ?

Bien sûr, nous ne voyons pas ce qui fonderait cette portée réelle : mais ne pas voir une chose n'autorise pas à la nier, cela autorise seulement à dire : la question demeure ouverte.

Oui ! la question demeure ouverte et nous ne pensons pas que la seule étude logique de la statistique soit capable d'y fournir une réponse : une telle réponse n'est pas du ressort de la logique, elle est du ressort de la philosophie - c'est-à-dire avant tout du bon sens (1)

Ce qui se produit ici est un phénomène psychologique remarquable : ce que nous connaissons du monde objectif n'est généralement pas exactement ce qui nous intéresse, ce qui nous intéresse est souvent quelque chose de futur, ce que nous connaissons est du passé ; de toute manière, ce qui nous intéresse est précisément ce que nous ne "tenons" pas encore (même si c'est quelque chose du passé) : si nous fixions notre regard sur ce que nous "tenons", nous aurions une connaissance directe. Mais notre condition humaine consiste au contraire à toujours poursuivre ce que nous ne tenons pas, et, sitôt que nous le tenons de quelque façon, à nous en faire un tremplin pour essayer d'atteindre quelque chose d'autre. Ainsi, ce que nous tenons ne nous intéresse que par les possibilités qu'il renferme de nous faire accéder à autre chose, et, de ce fait, nous avons devant le monde une attitude de "strabisme intellectuel" ; l'oeil de notre connaissance conserve l'empreinte d'un certain connu, tandis que l'oeil de notre volonté "louche" vers quelque inconnu ; l'angle qui sépare ces deux directions se traduit par une atténuation de la valeur informante de notre connaissance, atténuation que nous exprimons en utilisant le mot de vraisemblance au lieu de certitude ; la vraisemblance, c'est la quantité d'information que notre connaissance fournit à notre volonté.

L'atténuation dont nous venons de parler peut aller jusqu'à l'absence totale d'information sur ce qui nous intéresse ; il y a alors, entre connaissance et volonté, "orthogonalité" ou, comme on dit on statistique : indépendance.

C'est cet "angle" entre connaissance et volonté qui atténue la valeur informante de l'assomption statistique, non le seul fait qu'il s'agisse d'assomption ; car il serait inexact de penser que toute assomption présente nécessairement un degré de certitude imparfait. C'est par assomption que nous recevons le témoignage d'un autre homme, et il est tel cas où ce témoignage offre une parfaite garantie de crédibilité. En ce cas, notre connaissance de ce qu'il nous a appris est bien une certitude sans être pourtant une connaissance directe : le propre de l'assomption est donc de ne fournir qu'une connaissance indirecte ; et ce qui, parmi les assomptions distingue la connaissance statistique, c'est qu'elle est, non seulement indirecte, mais en outre imparfaitement certaine (2) : elle n'est que vraisemblable.

L'objet de la statistique, avons-nous dit, est de fonder sur des principes cohérents les décisions que nous prenons. Parmi ces décisions, nous en avons examiné un type particulier : l'assomption d'hypothèse, dans le cas où tout résultat exact est affecté de la même valeur subjective, du seul fait qu'il sera exact. Il n'en est pas toujours ainsi et nous devons maintenant dire un mot du cas plus général où les valeurs subjectives des différents résultats possibles sont inégales (mais encore supposées mesurables)

(1) Pourquoi, depuis qu'il y a des philosophes, faut-il périodiquement leur rappeler ce dernier point ? ...

(2) On notera que la connaissance directe est toujours, au contraire, parfaitement certaine.

C. - LE THÉORÈME DE BERNOULLI

Supposons que dans l'exemple précédent des deux pièces de monnaie j'aie fait un pari avec un camarade, pari en vertu duquel je gagnerai 100 francs s'il y a deux pièces côté face, et je perdrai 10 francs s'il y en a 1 ou 2 côté pile. Cette fois-ci il est bien clair que la décision à laquelle j'attacherai la plus grande valeur aléatoire est l'"assomption" X = 0.

D'une manière générale, ce type de problème se traitera aisément par la comparaison des espérances mathématiques, toutes les fois du moins où l'on pourra admettre de façon suffisamment approchée les postulats auxquels nous avons fait allusion dans le premier paragraphe de ce chapitre (1).

Il n'est pas toujours nécessaire d'ailleurs de connaitre exactement les valeurs subjectives, il suffit, en bien des cas, de connaitre une borne inférieure de telle d'entre ces valeurs (proportionnellement aux autres), pour être conduit à prendre la décision correspondante, pour "assumer" l'hypothèse en question.

Admettons par exemple que sur la table placée loin devant moi se trouvent non plus 2 mais 1.000 pièces de monnaie, et que j'aie été amené à penser (ceci est ici traité comme hors de discussion) que les vraisemblances de "face" et "pile" sont équivalentes pour chaque pièce et indépendantes d'une pièce aux autres ; ces conditions sont regardées comme réalisées notamment dans le cas où les pièces ont été lancées en l'air pour déterminer leurs positions. Dans certains problèmes pratiques, il m'est indifférent de savoir les positions individuelles de chaque pièce, il m'est même indifférent de connaitre le nombre exact de pièces tournées coté pile : il me suffirait de savoir si ce nombre x est suppérieur à 460 par exemple.

Or, en admettant comme applicables ici les principes partiellement explicités dans les pages qui précèdent, nous pouvons conclure ceci : tant que la valeur subjective de l'éventualité $x < 460$ n'est pas plus de 1.000 fois supérieure à celle de $x > 460$, nous devons miser sur $x > 460$, nous devons "décider" de nous placer dans cette hypothèse, nous devons autrement dit l'assumer".

Tel est selon nous le sens du théorème de BERNOULLI : il me dit que je dois assumer telle hypothèse tant que la valeur subjective de sa contradictoire n'est pas considérable, c'est-à-dire presque toujours. Le théorème de BERNOULLI nous enseigne notre devoir, sans nous garantir qu'en le faisant nous réussirons.

Cependant c'est un fait, un fait dont nous avons esquissé plus haut une explication psychologique, qu'en "obéissant" aux consignes du théorème de BERNOULLI on réussit généralement : que signifie cela ? Ce n'est pas ici le lieu de donner à cette question une réponse plus précise, qui relèverait de la métaphysique ; faisons seulement quelques remarques :

D'abord celle-ci : on ne réussit pas toujours, et même, selon le même théorème de BERNOULLI, on ne doit pas réussir toujours -entendons : on ne doit pas "assumer" l'hypothèse qu'on réussira toujours. Si

(1) Attirons l'attention sur le fait qu'il n'en sera pas toujours ainsi et que les mathématiciens sont exposés à souvent manquer sur ce point de "l'esprit de finesse" désirable.

l'on prend habituellement une garantie de 1.000 contre 1, comme dans l'exemple ci-dessus, on doit "assumer" un échec sur 1.000 "en moyenne". Mais on sait aussi qu'en fait, à moins d'un hasard, la proportion d'échecs ne sera pas non plus exactement de 1 pour 1.000. En sorte que toute "affirmation" du calcul des probabilités contient en elle-même sa négation, paradoxe fort connu qui s'explique si la statistique est d'abord une science de décisions et, de façon indirecte seulement, une connaissance.

Ensuite, notons qu'on ne réussit pas en fait dans la proportion que permettrait d'espérer le théorème de BERNOULLI ; mais chaque fois qu'on s'écarte un peu "trop" de cette proportion, on se dit : c'est que mes hypothèses n'étaient pas légitimes, il faut les remettre en question : c'est là le problème des "tests", que nous étudierons plus loin, et qui nous montre une des raisons pour lesquelles la statistique réussit généralement ; c'est que, là où elle échouerait, ce n'est pas elle que nous accusons, mais nous voyons dans son échec un signe (statistique encore) qu'il nous faut assumer sur le monde, une hypothèse autre que celle assumée jusque là. On pourrait dire en ce sens que la statistique réussit alors par définition.

Mais ce n'est pas une explication suffisante : si nous tenons tellement à ce que la statistique ait toujours le dernier mot, c'est que nous lui croyons une certaine valeur objective, et c'est que, même sans coups de pouce, elle réussit en fait assez souvent.

Il faut sans doute en conclure que le monde où nous vivons possède une certaine "cohérence" interne, en vertu de laquelle le connu présente avec l'inconnu des relations profondes dont les vraisemblances nous manifestent quelque chose, quelque chose de réel ; le connu est, en quelque mesure, signe de l'inconnu, en particulier le passé signe du futur (peut être parce que l'essence du monde est une essence de signe ? - Nous laisserons à la métaphysique le soin de répondre à cette question) ; la cohérence, que la statistique s'efforce de réaliser dans notre esprit, aurait ainsi son principe dans une cohérence objective du monde.

Peut être, dans ces conditions, la statistique mathématique devrait-elle réussir précisément dans la proportion prévue par la Loi des grands nombres et que, s'il en est parfois autrement en pratique, cela tient à ce mécanisme psychologique d'auto-suggestion dont nous avons déjà parlé, par lequel nous laissons indûment nos désirs peser sur notre appréciation des vraisemblances. Ainsi, la première qualité d'un statisticien est-elle la loyauté intellectuelle, cette rectitude, ce désintéressement dans la manière de juger qui, sans supprimer certes le jeu des préférences personnelles, les confine dans leur domaine légitime : celui des "valeurs subjectives". Il est même bien possible que pour un esprit pareillement droit, la formule

$$D_{ij}/ij/V_i\ \varpi_j$$

soit toujours valable, toutes les fois, s'entend, où les V_i et les ϖ_j sont eux-mêmes mesurables, restriction qui n'a plus rien à voir avec la loyauté du sujet mais tient à la nature de l'homme.

Il reste que, même pour un esprit loyal, les "assomptions" ne seront pas toujours vérifiées par la réalité et demeurent des décisions fondées sur une connaissance imparfaite.

D. - SOLUTION DU PROBLÈME DE L'ESTIMATION

Jusqu'ici nous avons supposé que les vraisemblances sur lesquelles portaient nos calculs nous étaient fournies par voie de connaissance directe Mais, en pratique, il arrive souvent que ces vraisemblances (ou plutôt leurs rapports) sont elles-mêmes le fruit d'une assomption (1) : nous sortons alors du domaine du probabiliste (où nous nous étions cantonnés jusqu'ici) pour entrer dans le domaine propre du statisticien.

Le calcul des probabilités ayant pour objet de nous dire quelles décisions nous devons prendre pour agir dans le monde, en fonction de certains schémas admis, le rôle de la statistique mathématique est précisément d'arrêter ces schémas, de décider lesquels d'entre eux doivent être assumés.

Le statisticien rencontre ainsi deux types de problèmes : problèmes d'estimation, problèmes de test : mais nous verrons que le second ne diffère du premier que par son aspect psychologique. Parlons d'abord du premier.

Selon nous, en accord avec les remarques diverses du premier chapitre, et selon les principes exposés dans le chapitre présent, un problème d'estimation n'a de sens qu'à la double condition suivante :

1°) le statisticien a, au moins confusément, quelque idée sur les lois de probabilité qu'il envisage d'appliquer au phénomène étudié, et attache à ces lois certaines "vraisemblances" a priori (fonctions de son expérience passées, etc...) ;

2°) le "client" fournit au statisticien, implicitement au moins, un certain optimum à réaliser, autrement dit certaines valeurs subjectives ;

dans ces conditions, le rôle du statisticien est d'évaluer les valeurs aléatoires relatives des diverses solutions qu'il envisage de donner au problème posé, et de fournir à son client la solution dont la valeur aléatoire est la plus grande.

Le problème n'a de solution que si les vraisemblances d'une part, les valeurs subjectives d'autre part, sont mesurables au moins approximativement ; on ne doit pas oublier cette condition, qui n'est pas toujours remplie ; il faut encore que la formule $D // V . \varpi$ soit applicable, c'est-à-dire que les valeurs aléatoires soient elles aussi mesurables : nous pensons qu'on peut ordinairement l'admettre conformément à une suggestion du paragraphe précédent.

Dès lors, le problème est aisé à mettre en équation ; il admet du reste autant de types de solutions qu'il existe de types de problèmes ; en voici quelques exemples :

Des billes rouges et noires sont tirées d'une urne U dont nous savons qu'elle a été choisie entre deux, A et B : nous nous demandons laquelle ? A contient 50 % de rouges, B en contient 25 %. A priori, nous tenons pour également vraisemblable l'utilisation de A ou de B : $V(A) \asymp V(B)$, et nous pouvons exprimer cela en disant que V (A) et V (B) sont entre elles comme

(1) En quelle mesure des vraisemblances dont les rapports sont fixés par assomption sont-elles encore des vraisemblances proprement dites ? Question difficile dont nous ajournons l'étude.

1 et 1 (1). Si nous nous intéressons à la seule exactitude du résultat, les valeurs subjectives des deux assomptions possibles sont égales. Dans ces conditions, leurs valeurs aléatoires sont aussi égales, et nous choisirons indifféremment l'une ou l'autre.

Mais maintenant on effectue un tirage de 5 billes et 2 d'entre elles sont rouges, 3 noires. Les vraisemblances après tirage sont alors, conformément au chapitre précédent, proportionnelles à : V (A) × 10 et V (B) × 5 (2) soit à 10 et 5. Les valeurs aléatoires, si elles sont mesurables, sont donc aussi entre elles comme 10 et 5, et nous devons assumer l'hypothèse qu'il s'agit de l'urne A.

Il n'en irait plus de même si l'urne B possédait soit une valeur subjective, soit une vraisemblance a priori double ou plus que double de A.

Si nous sommes dans le cas des probabilités continues, ayant à choisir entre une loi f (x) dx et une loi g (x) dx, un échantillon ξ sera regardé, non comme un point mais comme un volume infiniment petit dω dans l'espace de configuration : dω étant le même pour f et pour g s'élimine de la proportion, et nous choisirons de f et g la loi à laquelle nous attachons la plus grande valeur aléatoire :

$$V(f)\varpi(f)\, f(\xi) \quad \text{ou} \quad V(g)\varpi(g)\, g(\xi)$$

Si, au lieu de deux lois, nous en avons une infinité dépendant d'un paramètre m, et si les vraisemblances a priori et les valeurs subjectives sont distribuées de façon homogène sur l'axe des m, nous retrouvons la méthode de Fisher. Mais on voit que pour légitimer cette méthode nous sommes obligés d'admettre que le paramètre m est un certain paramètre bien déterminé, un "paramètre propre" du problème : l'espoir nourri par le grand statisticien anglais d'éliminer cette servitude est déçu, et il est déçu parce qu'il doit l'être : le choix du paramètre n'est aucunement indifférent : si la chose est obscure et contestée (bien qu'à notre avis incontestable .) en ce qui regarde les vraisemblances, elle est claire et hors de discussion pour ce qui est des valeurs : notre chapitre I a suffisamment attiré l'attention, pensons-nous, sur ce primat du client, seul fondé à décider l'optimum qu'il nous demande de réaliser, et qui donne au choix de tel ou tel paramètre un caractère privilégié.

Notons maintenant que, dans les applications industrielles, on ne sera presque jamais dans le cas où s'applique théoriquement la méthode de Fisher, même avec pondération V (m). ϖ (m), parce que ϖ est presque toujours fonction, non seulement de m, mais aussi de la valeur estimée m'. Montrons-le en traitant deux exemples importants. Comme nous avons expliqué que le produit V (m) f (ξ ,m) est une vraisemblance a postériori dont l'origine par échantillonnage n'importe guère, nous désignerons désormais simplement par φ (m) ce produit, représenté graphiquement par l'"estimatrice" ; V = φ (m).

1°) CAS "PARABOLIQUE".

Supposons un "statisticien conseil" S travaillant à son compte pour des clients C qui lui demandent des estimations de paramètres m, avec la clause suivante : ce qui intéresse les clients étant de réduire au minimum

(1) En effet, puisqu'il n'y a que deux vraisemblances en jeu, cette notation est cohérente sans qu'il soit nécessaire d'examiner les autres conditions pour qu'un ensemble soit mesurable.

(2) Sous la réserve déjà formulée page 74 note (2).

le dommage que leur cause une estimation inexacte des m, dommage bien souvent proportionnel au carré de l'écart, S recevra des honoraires déterminés, mais devra par la suite indemniser ses clients proportionnellement à l'écart de ses estimations m' avec les valeurs vraies m (lorsque ces valeurs vraies seront connues).

Un tel contrat n'est pas aussi chimérique qu'il semble ; d'abord il reflète exactement la situation où se trouve une compagnie d'assurances lorsqu'elle est son propre statisticien : en sous-évaluant les risques, elle s'expose à des déficits ; en les surévaluant, elle sera conduite à adopter des tarifs qui éloigneront la clientèle : on peut admettre qu'au voisinage de l'optimum, les écarts ont des coûts sensiblement quadratiques. Mais il y a plus : presque tout statisticien se trouve implicitement dans les conditions précédentes, parce que, s'il ne risque pas habituellement des indemnités à verser, il risque sa réputation : celle-ci est entamée par toute erreur notable de sa part (même s'il n'est aucunement fautif .), et il se trouve pratiquement obligé d'en tenir compte. Bien que la réputation soit une valeur particulièrement peu mesurable, on peut cependant étudier "qualitativement" l'incidence de cette considération par la formule parabolique.

Le problème se complique lorsque on doit envisager le risque d'erreur ruineuse, c'est-à-dire tenir compte d'écarts très grands dont le prix serait très supérieur à la somme des prix des écarts partiels : ce cas se rencontre, mais ici nous l'écartons, en supposant que les vraisemblances des écarts décroissent plus vite que ne croissent leurs prix, et qu'ainsi on peut traiter les valeurs comme additives.

Il est facile alors de voir que l'on doit choisir pour m' le point moyen de l'estimatrice

$$m' = \frac{\int \varphi(m) \cdot m \cdot dm}{\int \varphi(m) \quad dm}$$

c'est la fameuse formule de Bayes, dont nous voyons ainsi et la justification, et le champ d'application : ce champ est limité, mais cependant assez étendu, sous réserve de ce que nous dirons après avoir étudié le cas suivant ; pratiquement, la formule de Bayes répond à presque tous les besoins du "statisticien conseil", étant exclus seulement les cas de risques ruineux.

2°) CAS "SÉCURITÉ".

En réalité, les problèmes d'estimation ne se présentent pas ordinairement aux industriels de la façon que nous venons d'exposer ; c'est pourquoi nous distinguons, du "statisticien conseil" avec qui un client étranger a passé un contrat, le "statisticien associé" qui identifie ses préoccupations à celles de la maison pour laquelle il travaille (soit parce qu'il est scrupuleux, soit parce qu'il est rétribué par participation proportionnelle aux bénéfices - et aussi aux déficits ?...

Pour l'industriel, la valeur m' adoptée pour m sera plus ou moins irrévocable : l'écart m' - m sera fixé une fois pour toutes, et son coût ira se répétant d'année en année : ce qui importe ici c'est que ce coût ne soit pas excessif. Pratiquement, on prendra ses précautions pour tenir compte de l'écart inévitable entre m' et la vraie valeur m, c'est-à-dire qu'on prévoira une marge de sécurité $m' - \alpha$, $m' + \beta$; on peut même dire que le nombre m' lui-même est généralement sans intérêt : ce qui compte, c'est le couple $m_1 = m' - \alpha$, $m_2 = m' + \beta$. Or, il arrive souvent qu'on

puisse choisir le paramètre m de telle sorte que le coût de la sécurité (m_1, m_2) soit fonction de la seule différence $m_2 - m_1 = \alpha + \beta = \Upsilon$; (fonction croissante, évidemment) : un tel paramètre est alors le paramètre propre du problème, et l'on trouve sans difficulté que le minimum de Υ pour une vraisemblance relative donnée - ou le maximum de cette vraisemblance pour un Υ donné, s'obtient en plaçant Υ sur l'estimatrice de manière que les ordonnées limites $\varphi(m_1)$ et $\varphi(m_2)$ soient égales : c'est la méthode de Neyman, dont nous voyons ici et la justification, et le champ d'application.

On notera que si l'on s'intéresse à m' lui-même ce sera généralement le milieu de l'intervalle (m_1, m_2) qu'il faudra lui attribuer.

REMARQUES :

1°) La comparaison de ces deux cas est extrêmement révélatrice : elle tend à démontrer mathématiquement cette vérité de bon sens, que pour obtenir d'un homme qu'il épouse les intérêts de la maison qui l'emploie, il ne suffit pas de stimuler son travail par des primes au rendement ou des pénalités ; il faut, s'il n'est pas d'un désintéressement angélique, que sa "fonction prix" soit homothétique à celle de la maison elle-même : ce qui signifie en bon français qu'il ne soit pas traité en salarié, mais en associé.

Notre démonstration n'a été faite que pour le métier de statisticien : mais tout homme qui réfléchit tant soit peu avant d'agir est un statisticien implicite, comme M. Jourdain un prosateur.

Faisons remarquer cependant que le problème qui se pose au statisticien de sauvegarder sa réputation crée une sérieuse difficulté. Cette difficulté ne disparaitra que le jour où les hommes auront une culture statistique suffisante pour comprendre les risques qu'un bon statisticien doit accepter de courir pour s'acquitter honnêtement de sa fonction, faisant passer l'intérêt de ses clients avant le soin de sa réputation personnelle. C'est aussi, bien entendu, une question de confiance.

2°) Une seconde remarque va cependant beaucoup minimiser la portée de la précédente : dès que l'on dispose, pour estimer un paramètre m, d'un échantillon un peu important, l'estimatrice $V = \varphi(m)$ se rapproche rapidement d'une courbe en cloche, en sorte que même si elle possède encore une dispersion notable, sa symétrie approchée suffit à faire coïncider pratiquement les résultats des divers procédés d'estimation, <u>du moins de tous ceux qui sont des fonctionnelles de l'estimatrice</u> : telle est la véritable portée de la découverte de Fisher ; ce n'est pas le "maximum de vraisemblance" qui est intéressant (encore que ce soit généralement le point le plus <u>commode</u> à calculer, en sorte que c'est presque toujours lui qu'on continuera à utiliser) ; mais ce qui est essentiel c'est que, pour les ajustements adéquats, <u>l'estimatrice est une "caractérielle exhaustive"</u> car V (m) joue également très peu : tous les statisticiens sont assez vite également "bons").

3°) Un mot sur le problème des "ajustements inadéquats" : plaçons-nous dans le cas où l'estimation se fait par la formule de Bayes, cas pratiquement assez général en vertu de la remarque ci-dessus. Alors, si nous cherchons "la" loi correcte (1) dans la famille f (x,m) nous ne la trouverons pas, puisqu'elle ne s'y trouve pas. On peut alors imaginer que la famille f(x,m) est "plongée" dans une famille plus large $f(x,m,\lambda)$; et bien

(1) Ecartons ici le problème (philosophique) de savoir quel sens a cette façon de parler, et si même elle en a vraiment un.

entendu on peut toujours supposer que cette seconde famille contient la vraie loi. L'estimation de m se ferait alors par la formule :

$$m' = \frac{\iint f\, \varphi\,(m)\ \psi\ (m, \lambda)\ m\ dm\ d\lambda}{\iint f\ \varphi\ \psi\ dm\ d\lambda}$$

ou encore ;

$$m' = \frac{\int \varphi \cdot m\ \cdot\ dm \cdot \int f\ \psi\ d\lambda}{\int \varphi \cdot dm \cdot \int f\ \psi\ d\lambda}$$

ce qui peut s'écrire :

$$m' = \frac{\int \theta\,(m)\ g\ (x,\ m)\ m\ dm}{\int \theta \cdot g\ \cdot\ dm}$$

Il n'est donc pas nécessaire, pour estimer m, que l'on connaisse une classe f (x,m) contenant la "vraie" loi, une classe "adéquate" : il suffit d'avoir à sa disposition une classe g(x,m) qui soit seulement "adéquate en moyenne". On notera que g peut fort bien ne pas appartenir à la classe $f(x,m,\lambda)$; c'est donc à la recherche d'une classe adéquate en moyenne, non à la chimère d'une classe contenant réellement la vraie loi, que le statisticien est ainsi ramené.

4°) Il faut noter que la détermination des valeurs subjectives, des prix qui sont en jeu dans un problème d'estimation, est une tâche qui normalement reviendrait au "client", puisque c'est lui qui connait le but qu'il poursuit, lui qui sait la valeur qu'il attribue à telle ou telle éventualité : mais en pratique le statisticien sera d'ordinaire seul compétent pour exprimer ces valeurs en formules mathématiques - à supposer, évidemment, qu'elles forment un ensemble mesurable.

Cette dernière condition ne va aucunement de soi ; disons même qu'il sera bien rare de se trouver en présence de valeurs exactement mesurables, et l'on devra se contenter de mensurabilité approchée. Mais celle-là même sera-t-elle toujours admissible ? On devra s'attendre à des surprises, et il ne sera pas toujours facile de faire comprendre au "client" que son désir échappe à ce que les mathématiques peuvent mettre en équation. En ce cas on cherchera un terrain d'entente, en prenant le problème par quelque biais, et substituant un optimum calculable à celui qui ne l'était pas.

E. LE PROBLÈME DES TESTS

On sait en quoi ce problème consiste : à un moment donné on a admis qu'un certain phénomène était représentable par un schéma H de probabilités ; par la suite, après avoir recueilli de nouvelles informations, ou bien encore (ce qui semble revenir au même dans une certaine mesure) après avoir porté son attention sur un nouvel aspect des informations anciennes, on doute de la validité du schéma H : quelle règle nous permettra de décider si l'on doit conserver H ou non ?

La solution de ce problème dépend de deux remarques : d'abord qu'en admettant le schéma H on a fait une "assomption" d'hypothèse ; l'a-t-on faite à bon droit, cela n'est pas en question ; il s'agit de savoir si on doit maintenir cette assomption, compte tenu des éléments de décision plus abondants que nous possédons maintenant.

La seconde remarque, c'est qu'en assumant H on envisageait au moins implicitement la possibilité d'autres assomptions (sinon il n'y aurait pas eu assomption : la certitude seule exclut l'arrière-pensée de possibilités différentes) : tester, c'est revenir sur ces autres possibilités, et les explicitant partiellement, comparer leurs valeurs aléatoires actuelles à celle de H.

Ce qui montre bien qu'il en est ainsi, c'est que, quel que soit le test employé, lorsque ce test aura conclu au rejet de H, nous remplacerons H par une autre hypothèse K ; c'est donc que nous avons K en vue, et puisque c'est notre test qui nous conduira éventuellement à essayer K, c'est qu'il existe une certaine affinité entre ce test et K ; un test bien fait doit être orienté vers l'hypothèse K par laquelle on envisage de remplacer éventuellement H : sinon, c'est que ou le test, ou K sont mal choisis. Lorsque, par exemple, on teste par la moyenne expérimentale la correction d'une loi de Gauss dont le centre est à l'origine c'est qu'implicitement on envisage de déplacer ce dernier.

Dès lors, tester H n'est pas autre chose que de faire une estimation dans la classe H + K, estimation dans laquelle, si H a été correctement choisi, la valeur aléatoire de H avant considération des données nouvelles devrait être supérieure à celle de K, mais a pu cesser de l'être après.

Si H et K sont par exemple des lois de probabilité f (x) dx et g (x) dx, et si les données nouvelles consistent en un échantillon ξ , on sera conduit à comparer :

$$\varpi(f) \cdot V(f) \,.\, f(\xi) \text{ à } \varpi(g) \,.\, V(g) \,.\, g(\xi) \quad (1)$$

Il sera parfois difficile de fixer les valeurs ϖ V ; mais on procèdera alors ainsi : f doit être conserver si $g(\xi)/f(\xi)$ n'est pas supérieur à

$$\lambda = \varpi(f)\, V(f) / \varpi(g)\, V(g)$$

Or, si λ est mal défini (et même souvent peu mesurable), on peut habituellement en donner une borne inférieure : si $g(\xi)/f(\xi)$ (2) est plus petit que cette borne, on doit conserver f ; ou bien encore : on calcule d'abord $g(\xi)/f(\xi)$, et l'on se demande si λ peut être plus grand, ou au contraire plus petit.

Notre test est donc le "likelihood ratio" de Neyman et Egon Pearson. Mais on remarquera que nous l'utilisons autrement : nous ne nous intéressons plus à la notion "limite de confiance"; c'est le rapport même g/f des densités qui est utilisé, et directement comparé à λ, sans aucune préoccupation de la probabilité de : $g/f > \lambda$. En fait, cela ne produira que de faibles changements dans les tests où H et K diffèrent par un paramètre : dans ces cas là, à la limite de confiance 1 % sera simplement substituée la valeur relative $\lambda = 20$ environ, etc... Mais il en ira autrement avec une différence supérieure à l'unité dans le nombre de paramètres : conserver $\lambda = 20$ se traduira par des probabilités d'autant inférieures à 1 % que cette différence sera plus grande.

Il y a un cas qui soulève quelque difficulté : c'est celui où l'on envisage, après application du test, au lieu de remplacer immédiatement H

(1) On notera que dans un problème de connaissance scientifique désintéressée les ϖ doivent être pris égaux.

(2) Ou la borne supérieure de g/f lorsque g contient un paramètre.

par une autre hypothèse, de procéder auparavant à de nouvelles épreuves : et c'est seulement sur la vue de ces dernières qu'on choisira K. Alors nous croyons qu'on doit tester ainsi : si nous ne sommes pas en état de fixer K d'avance, nous avons cependant quelque idée du genre d'hypothèse que nous ferons à son sujet ; et cela suffit pour évaluer grossièrement une borne supérieure de $g(\xi)/f(\xi)$: si cette borne supérieure est plus petite que la borne inférieure de λ (1), nous n'irons pas plus loin et garderons H.

On voit que dans de pareilles situations (qui sont des plus fréquentes en pratique), il faut renverser une formule classique et dire : la Nature, en ces cas, ne peut répondre que oui. Seulement son oui ne signifie pas : les choses sont bien ainsi ; il signifie : vous avez le devoir de demeurer en repos et d'admettre que les choses sont ainsi (encore que ce soit peut être inexact).

Au contraire, il sera souvent plus délicat qu'on ne pense de décider le rejet de H ; rarement ce rejet se fera de façon définitive sur la simple vue d'un test : on voudra joindre à cette donnée brutalement numérique une certaine compréhension de la nature des faits en cause et du motif de l'échec subi. Et c'est fort bien ; car il ne convient pas de réduire le statisticien au rôle de machine à calculer.

NOTE A. - LA SIGNIFICATION DES PROBABILITÉS ESTIMÉES

Nous nous sommes demandés plus haut (note 1 p. 79) si l'on pouvait encore regarder comme quotients de vraisemblances les probabilités résultant d'une "assomption". Nous pensons que la réponse est affirmative, eu égard aux remarques suivantes :

Appelons S un "statisticien", et désignons par S_1 l'état où il se trouve au moment où il va entreprendre une estimation - mettons : composition (p, 1 - p) d'une urne, connaissant un échantillon (k, n - k). S_1 dispose des données suivantes :

Le système Σ de ses principes intellectuels, et notamment de ses principes statistiques ; un système (V) de vraisemblances a priori touchant la composition de l'urne ; l'échantillon (k, n - k) lui-même ; enfin un problème pratique à résoudre donnant lieu à un système de prix.

Ces données n'étant pas utilisables telles quelles pour l'action, S prend d'emblée une première décision, fait une première assomption (à laquelle on ne songe pas toujours assez) : celle d'attribuer pour l'avenir (et ceci déborde le problème immédiat donnant lieu aux prix ϖ), d'attribuer donc à l'urne et aux résultats des futurs tirages qu'on en fera un schéma probabiliste (p, 1 - p). Dès lors, S n'est plus dans l'état S_1 , mais dans un nouvel état S_2 ne disposant plus, au lieu des (V) et de la composition précise de l'échantillon (comportant même peut être l'ordre de sortie des boules ...) que de la seule estimatrice :

$$v = \varphi(p) = V(p) \quad p^k(1-p)^{n-k}$$

Le passage de S_1 à S_2 constitue, disons-nous, une première assomption (plus ou moins implicite) : celle-ci suppose un système (également implicite) de valeurs, en vertu duquel on estime avantageux de substituer aux connaissances directes mais inutilisables de S_1 , le schéma assumé par S_2. Ce système de valeurs appartient du reste à Σ : c'est dès le début de notre vie humaine que nous avons décidé de substituer ainsi, aux données concrètes de l'espérience, difficilement maniables dans leur richesse et leur nuance, des schémas simplifiés intérieurs à notre esprit (images, mots ...), constituant une sorte de microcosme, ou "intramonde", nécessaire pour l'action. Nous avons discuté précédemment de la valeur objective de cet intramonde, substitut du monde réel, auquel il possède du reste une certaine adéquation (à la philosophie de préciser).

Ensuite, une seconde assomption, comme on sait, mène S de S_2 à un nouvel état S_3, dans lequel l'estimatrice $v = \varphi$ (p) est réduite à un seul point p' en considération des ϖ.

Qu'est donc p' pour S ? Il faut ici bien remarquer que S est maintenant devenu S_3 : de S_1 à S_2, puis à S_3, il y a plus qu'un passage logique : il y a modification réelle. S_3 est désormais le seul état concret de S, et S_1 n'est plus qu'un souvenir. En effectuant les deux assomptions que nous avons étudiées, en prenant les deux décisions susdites S a cessé d'être S_1 il est devenu S_2, puis S_3. Et si par la mémoire S_3 peut bien se rappeler que, pour S_1, p' faisait figure d'hypothèse parmi d'autres possibles, pour lui maintenant p' fait partie de l'image qu'il attribue au monde. Que ce soit à tort ou à raison, là n'est pas la question : il sait bien en effet (ou devrait savoir) que cette image, résultant en partie d'assomptions, est d'une exac-

(1) Et dans λ cette fois on doit faire entrer commercialement parlant le coût des épreuves supplémentaires.

titude peut être imparfaite, et certainement discutable ; néanmoins S est maintenant "installé" dans cette image du monde, et provisoirement il y demeurera.

Dès lors S_3 est "quelqu'un" d'ainsi constitué qu'il attribue aux sorties des diverses couleurs de boules des vraisemblances qui sont dans le rapport p'. Quand aux valeurs absolues des vraisemblances, elles ne sont pas suffisamment déterminées par ce qui précède, mais résulteraient (si l'on en avait besoin) des autres informations de S dont nous n'avons pas eu à faire état, et qui se retrouvent dans S_3 comme dans S_1 (avec les seules modifications qu'y ont introduites les assomptions par lesquelles S_3 diffère de S_1).

Remarque

Nous avons étudié dans notre dernier chapitre, paragraphe E, un aspect du problème des tests qui n'est pas essentiellement différent d'un problème d'estimation : arrivés à l'état S_3 en vertu d'assomptions antérieures, de nouvelles informations changeaient S_3 en S'_3 , et de là, par assomption, nous passions à quelque S_4 . . . (1)

Il existerait un autre aspect qui, au contraire, consisterait à remonter de S_3 vers S_1 (et de proche en proche toujours plus haut) afin de rechercher dans notre esprit ce qui s'y trouve de connaissance directe du monde, non simplifiée (et peut être déformée) par des assomptions utilitaires : c'est là le "problème critique" des philosophes, où l'on s'efforce de saisir la réalité telle qu'elle EST, et non point telle que nous l'UTILISONS

Nous pensons que c'est en remontant maille par maille la chaine, donc en partant du point où nous étions arrivés, qu'on a chance d'aboutir et non point, comme le voulait Descartes, en faisant table rase et recommençant à zéro : à zéro il n'y aurait plus que l'esprit d'un nouveau-né : pur de tout pragmatisme, soit : mais peu capable apparemment de philosopher . .

Tandis que notre "intramonde", s'il est quelque peu souillé de pragmatisme, contient néanmoins de la connaissance réelle : au philosophe de dégager celle-ci en "décapant" la croûte verbale et imaginative qui la recouvre.

Mais on voit qu'une saine philosophie doit procéder exactement en sens contraire du pragmatisme scientifique, et que rien n'est plus scandaleux que d'avoir prétendu fonder un jour la métaphysique sur une technique d'"assomptions" : prolégomènes à tout ce que l'on voudra, sauf précisément à de la métaphysique ! . . .

Il est inévitable qu'en toute métaphysique "faite de main d'homme" subsiste une part d'assomption, particulièrement dans le vocabulaire : mais c'en est la partie fragile, sujette à révision, la partie la moins "métaphysique" : il était réservé à certaine Ecole philosophique d'ériger en idéal ce qui n'est qu'une servitude de l'humaine faiblesse. . . Pourquoi nous sommes-nous laissés prendre à cette duperie ?

---

(1) Note de l'auteur en marge du manuscrit : "Ce processus n'est pas tout à fait exact."

NOTE B

Notre conception de la statistique parait soulever une sérieuse objection que l'on n'a pas manqué de nous faire :

Elle est tellement relative au statisticien qui en fait usage, que l'on risque de ne pouvoir s'accorder d'un statisticien à un autre. En d'autres termes, la statistique telle que nous la considérons serait strictement subjective.

Pour répondre à cette objection, nous demandons à soigneusement distinguer le problème pratique et le problème théorique.

1°) En pratique, la difficulté en question n'existera guère : dans la plupart des cas, les divers statisticiens seront d'accord pour attribuer aux vraisemblances a priori des valeurs peu différentes de l'un à l'autre, et généralement en outre extrêmement "aplaties" ; dans ces conditions, les estimations de ces divers statisticiens seront très vite pratiquement confondues.

S'il arrive qu'il n'en soit pas ainsi, on sera conduit à examiner les différentes estimations correspondant au "champ" des diverses vraisemblances a priori susceptibles d'être envisagées ; et l'on utilisera une méthode d'estimation tablant sur le risque maximum dans un tel champ (méthode de Wald, etc...)

2°) En théorie, le caractère subjectif (nous aimerions mieux dire : relatif) de la méthode subsiste ; mais c'est que, croyons-nous, il doit subsister parce qu'il est dans la nature des choses. Nous ne croyons pas aux notions de probabilités et de vraisemblances en dehors des hommes pour qui ces probabilités et ces vraisemblances existent. Lorsque je vois une fumée, elle provient d'un feu ou d'un dispositif fumigène, ou de quelque autre cause ; objectivement, l'une est vraie, les autres fausses, mais pour moi chacune à sa vraisemblance. De même, lorsque je lance une pièce en l'air, elle tombera en fait ou sur pile, ou sur face ; c'est seulement pour celui qui ignore le résultat (1) que pile et face ont chacune sa probabilité (que l'ignorance vienne de ce que la pièce n'est pas encore tombée, ou de ce qu'elle est trop loin ou cachée).

Il n'y a pas de probabilités objectives, mais seulement (comme nous l'avons expliqué au 1°) des probabilités pratiquement intersubjectives.

(1) Ou plus précisément : qui n'a que des informations imparfaites le concernant ; car nous n'admettons pas de probabilités fondées sur une pure ignorance.

NOTE C

En appendice à nos réflexions sur la notion de valeur, voici sur les valeurs économiques quelques considérations dont nous sommes, pour une large part, redevable à des échanges de vues avec M. MORLAT :

La valeur telle que nous l'avons définie est ce que les économistes nomment "valeur d'usage" : nous avons attiré l'attention sur son caractère relatif : relatif à l'usager, et plus précisément à cet usager se trouvant dans une certaine situation bien déterminée, avec des moyens déterminés, et une façon déterminée de juger de ses divers besoins ou aspirations. Cet usager, du reste, peut être une personne physique ou une personne morale.

Lorsque deux personnes (physiques ou morales) envisagent de réaliser librement un échange de biens (objets matériels, services, argent, etc...), cet échange n'a lieu que dans la mesure où chaque partie le trouve avantageux. L'échange n'est donc possible, sauf contrainte, que si les valeurs relatives des biens échangés sont différentes d'une personne à l'autre : cela suffirait à montrer que ces valeurs ne sont aucunement proportionnelles aux prix objectivement attachés aux biens dans nos sociétés modernes, sans quoi il n'y aurait presque jamais d'échanges.

En fait, un marché libre entre deux personnes se réalise par la constatation qu'une certaine proportion dans l'échange d'un bien A et d'un bien B est intermédiaire entre deux points limites : le point où le marché cesse d'intéresser A ; celui où il cesse d'intéresser B. L'audace d'une partie jointe à la timidité de l'autre tend à rapprocher la proportion de l'un ou l'autre point limite : une commune loyauté tendrait à la situer dans une zone médiane "également avantageuse" pour toutes deux. (1)

Mais la plupart des marchés sont d'un type différent, qui fait intervenir les valeurs subjectives non de deux, mais d'un grand nombre de personnes. Ces marchés ne sont donc plus des contrats privés, mais des contrats publics ; et il faut entendre par ces deux expressions deux réalités juridiques essentiellement différentes : un grand pas sera franchi dans l'organisation de la vie sociale le jour où l'on aura compris cette différence spécifique.

Le contrat privé est ordinairement bipartite ; le contrat public est, de sa nature, tripartite au moins ; il lie, explicitement ou implicitement, les deux parties qui échangent et une troisième partie qui est la communauté nationale ou internationale. Celle-ci a donc son mot à dire dans tout contrat public, puisqu'elle y est d'office partie (même si pratiquement elle n'a pas jusque là fait état de son droit) : c'est le cas de tout industrie ou commerce vendant avec publicité (l'enseigne commerciale est une véritable publicité) ; c'est le cas de toute entreprise dont les dimensions excèdent les besoins ou les moyens d'exploitation d'un homme seul (avec sa famille).

Nous pensons que cette notion de contrat tripartite fournirait des solutions nouvelles aux grands problèmes sociaux et économiques de l'heure, solutions qui échapperaient au dilemne : liberté quasi-totale ou nationalisations. Le dirigisme de la période des restrictions n'a certainement pas été une réussite parfaite. Mais c'était peut être la première tentative assez grossière d'une organisation qui demanderait à être repensée.

Quoi qu'il en soit de ces dernières considérations, dès qu'un bien est mis en assez large circulation dans une société, l'ensemble des fac-

(1) Ceci a un certain sens, malgré le caractère subjectif des valeurs : car l'homme possède, s'il le veut, une certaine aptitude à "se mettre à la place" d'un autre homme.

teurs qui entrent en jeu dans l'échange de ce bien contre d'autres (valeurs subjectives du bien en question pour les diverses personnes physiques et morales membres de la Société, abondance de ce bien, règlementation de son marché, etc...) a une résultante. Dans un système social où tout bien est censé échangeable contre un bien étalon unique par la convention du signe monétaire, la résultante en question se traduit par un prix.

Le prix, assez indûment appelé par les économistes "valeur d'échange" n'est pas de la nature d'une valeur proprement dite. D'une part il est une résultante complexe de facteurs qui sont en partie (mais pas tous) des valeurs. D'autre part, une fois constitué, il apparait comme un paramètre de structure du système économique : celui qui veut aujourd'hui échanger des signes monétaires contre du pain sait dans quelle proportion cet échange est possible : à lui de juger si la valeur qu'a pour lui un kilo de pain est supérieure ou inférieure à la valeur qu'ont, pour lui toujours, sept pièces de cinq francs (ces dernières étant échangeables aussi contre une heure de "Cinéac", etc...) Le prix n'est pas une valeur : il constitue une "équation de liaison" à adjoindre à tout problème de décision.

Pour compléter ces remarques, esquissons une classification des principales "Sciences des valeurs" :

L'étude des "valeurs subjectives" relèverait en principe de la psychologie ; c'est au probabiliste que revient normalement celle des vraisemblances, et au statisticien (en liaison, bien entendu, avec le probabiliste et le psychologue d'une part, avec l'usager d'autre part) celle des "valeurs aléatoires".

Quant aux "valeurs d'échange" (pour conserver un terme consacré), leur étude appartient à la sociologie (utilisant les disciplines précédentes): celle-ci se confond du reste avec la morale, car on ne peut étudier une société humaine comme une société animale : il y faut une référence à certains principes directeurs, une sorte d'"étalon" des valeurs humaines qui constitue précisément la morale. Quant à ce que l'on nomme parfois "morale personnelle", celle-ci n'est au fond qu'un chapitre tronqué de la théologie, une théologie sans Dieu.

La théologie, enfin, étudie l'homme dans sa relation à Dieu, l'"homo religiosus", l'homme envisagé comme une "personne" conditionnée par sa psychologie individuelle et son appartenance à la communauté humaine, mais transcendante à l'une et à l'autre.

Est-il utile d'ajouter que la science économique n'est elle-même qu'un chapitre, particulièrement complexe certes de nos jours, de la sociologie (ou morale) dont elle est inséparable ? Oublier cela, c'est ne plus savoir si les machines tournent pour nourrir les hommes, ou si les hommes mangent pour faire tourner les machines : il n'est que trop vrai que nous en sommes là ...

NOTE D

Voici d'autre part une très intéressante remarque due à M. LE CAM: dans mon étude des "grandeur psychologiques" j'ai presque d'emblée admis que la comparaison de celles-ci se fondait uniquement sur une relation soit d'équivalence, soit de prévalence : qu'il y ait des "degrés" dans la prévalence, cela résulte pour moi de la structure d'ensembles ordonnés au moyen d'un prévalence simple.

Mais de même que pour une surface, en outre de la distance géodésique de deux points (laquelle résulte de proche en proche de la seule donnée d'une distance élémentaire), on peut encore se donner une distance spatiale (la longueur de la corde) et que cet seconde donnée ne se réduit pas à la première et fournit à la surface une structure plus complexe : de même, M. LE CAM pense qu'il existe en psychologie entre deux grandeurs A et B une prévalence douée de degré avant toute comparaison de A et de B avec les autres grandeurs. Cette idée parait très intéressante. Comme elle introduit dans la question une énorme complexité, il me semblerait raisonnable d'attendre, pour étudier l'axiomatique correspondante, d'avoir auparavant complètement mis au point celle que je propose ici ; outre qu'elle en est un cas particulier très important en pratique, je pense que sa connaissance sera nécessaire avant de passer au degré suivant de complexité ; il faut en effet pouvoir "raccorder" la métrique spatiale à la métrique géodésique, et pour cela commencer par connaitre celle-ci.

CONCLUSION

La statistique n'a pas pour objet direct de nous dire ce qui est : dans les cas où ce que nous connaissons déjà ne nous intéresse pas en soi-même, ou au contraire ce qui nous intéresse ne nous est pas encore connu, elle nous enseigne avant tout ce que nous devons décider. Parmi ces décisions se trouvent notamment les assomptions d'hypothèses, qui suppléent à notre manque de connaissance directe sur ce qui nous intéresse, et nous fournissent à cet égard un schéma dont la fonction est double : directement ce schéma règlera notre conduite ; indirectement, il nous donne de ce qui est quelque idée plus ou moins approchée et plus ou moins certaine.

Car, en fait, notre schéma statistique est généralement efficace, et cette efficacité habituelle nous est un signe que le monde présente une certaine cohérence, et notre esprit une certaine adéquation au monde. Toutefois, ce schéma se ressent de son origine utilitaire : il est beaucoup plus pauvre que le très complexe faisceau de vraisemblances qui lui a donné naissance, il a délibérément sacrifié une part considérable de nos informations sur le monde.

Et voilà peut être pourquoi, utilitaires (donc statisticiens instinctifs) dès notre première enfance, nous nous sommes bâti une image du monde faite des "observables" α de Dirac, réduites encore à une seule valeur propre chacune ; nous n'avons pas voulu voir (parce que nous n'aurions rien pu en tirer pour notre vie pratique), que les observables sont des grandeurs, réelles sans doute, mais relatives ("objectivement relatives"), et que, en dehors de sa relation à nous, le monde est constitué par les "états" ψ : de ces états nous ne pouvons malheureusement pas avoir une connaissance particulière, parce que notre liberté, en nous rendant partiellement autonomes vis-à-vis du monde matériel, met entre les ψ et nous la marge du principe d'incertitude.

Ainsi nous trouverons-nous, lorsqu'une cosmologie nouvelle nous aura livré le secret des ψ, devant une double connaissance du monde : l'une absolue mais abstraite, fondée sur les ψ ; l'autre relative, mais concrète, fondée sur les α. Le savant préfèrera la première, le poète restera comme par le passé attaché à la seconde ; un philosophe digne de ce nom devrait savoir associer les deux ; car les deux sont réelles, les deux sont "de l'être" : si le monde existe, et si j'existe, les relations du monde à moi sont "quelque chose" du monde, tout comme les relations de moi au monde sont "quelque chose" de moi.

ÉNERGIE CINÉTIQUE MOYENNE D'UNE PARTICULE EN MOUVEMENT BROWNIEN DANS UN FLUIDE RÉEL

par

J. L. SOULÉ

Résumé - Cette énergie cinétique est moindre que celle qui résulte du principe d'équipartition. On l'évalue moyennant certaines hypothèses, valables pour un rayon de particule assez grand.

INTRODUCTION

A. EINSTEIN a publié en 1905 (1) une étude théorique quantitative du mouvement brownien de translation d'une particule sphérique dans un fluide, aboutissant à la célèbre formule suivante, qui a fait l'objet de nombreuses vérifications expérimentales :

(1) $$2\,D = \lim_{t \to \infty} \left(\frac{\overline{x^2}}{t}\right) = \frac{RT}{N}\,\frac{1}{3\pi\eta r}$$

D étant la constante de diffusion brownienne, $\overline{x^2}$ le carré moyen du déplacement de la particule selon une direction pendant le laps de temps t, R la constante des gaz parfaits, T la température absolue, N le nombre d'Avogadro, η la viscosité du fluide et r le rayon de la particule, Einstein généralisa et simplifia sa démonstration dans son mémoire suivant sur le sujet (2)

Une démonstration différente fut d'abord donnée de façon approximative par SMOLUCHOWSKI (3), puis reprise par Einstein lui-même (4), enfin mise sous une forme définitive très simple par LANGEVIN (5)

Cette seconde démonstration est souvent considérée comme plus satisfaisante que la première. En fait, nous nous proposons de montrer qu'elle n'aboutit au résultat exact que par un intermédiaire incorrect - sauf lorsque la masse spécifique du fluide est négligeable par rapport à celle de la particule ;

Notre étude ne concernera que le mouvement brownien de translation mais ses principes s'appliqueraient au mouvement brownien de rotation.

RAPPEL DE LA DÉMONSTRATION DE LANGEVIN

Rappelons rapidement les termes de la démonstration de LANGEVIN. Tout d'abord on évalue le carré moyen de la vitesse de la particule selon une direction, grâce au principe d'équipartition de l'énergie :

$$m\,\overline{u^2} = \frac{RT}{N} \tag{2}$$

Ensuite on relie $\overline{u^2}$ à $\lim \left(\frac{\overline{x^2}}{t}\right)$ par une étude hydrodynamique du mouvement. A cet effet, on distingue dans les impulsions communiquées à la particule par les molécules du fluide qui la heurtent deux parties : d'une part, une valeur moyenne (c'est-à-dire en fait une espérance mathématique) qui s'indentifie avec les composantes de la pression hydrodynamique et qui se relie au mouvement de la particule par la formule de Stokes ; d'autre part, une succession d'effets aléatoires indépendants L'équation du mouvement s'écrira :

$$m\frac{d^2x}{dt^2} + 6\pi\eta r\frac{dx}{dt} = F(t) \tag{3}$$

où F (t) est la part aléatoire de la pression.

En multipliant par x et en intégrant de 0 a t, on obtient ;

$$m\,x\,u(t) - m\int_0^t u^2\,dt + 3\pi\eta r x^2 - \int_0^t x\,F\,dt = 0 \tag{4}$$

En prenant les espérances mathématiques de chacun des termes, il vient :

$$m\,\overline{x\cdot u(t)} - m\,\overline{u^2}\cdot t + 3\pi\eta r\,\overline{x^2} - \int_0^t \overline{x\cdot F}\,dt \quad 0 \tag{5}$$

Il est intuitif que le premier et le quatrième termes deviennent négligeables à côté des 2 autres lorsque t augmente indéfiniment (voir par exemple pour une justification rigoureuse : (6)), et on a l'égalité asymptotique :

$$\lim \frac{\overline{x^2}}{t} = \frac{m\,\overline{u^2}}{3\pi\eta r} \tag{6}$$

qui établit, par comparaison avec (2), la formule (1).

Les variantes de cette démonstration ne sont différentes que dans la manière de passer de l'équation (3) à la formule (6).

DEUX OBJECTIONS A LA DÉMONSTRATION DE LANGEVIN

Si l'on tient compte de l'inertie du fluide, on doit faire une double objection à la démonstration que nous venons d'exposer :

1) La formule de Stokes n'est valable que pour un mouvement uniforme (nous ne parlerons pas de la correction de "glissement" ou correction de Cunningham, à faire intervenir lorsque r n'est pas assez grand devant le

libre parcours moyen des molécules du fluide). Puisque la vitesse de la particule est constamment et rapidement variable, l'inertie du fluide a pour conséquence que la résistance hydrodynamique qu'il oppose au mouvement de la particule dépend à chaque instant de toutes les accélérations antérieures, et en fait conduira à une vitesse moins fluctuante que ne le voudrait la loi de Stokes. Par suite, pour une vitesse quadratique moyenne donnée, le déplacement quadratique moyen sera plus grand que celui qu'on calcule avec la loi de Stokes.

2) Comme l'indique Einstein (4), lorsqu'on applique le principe d'équipartition, on suppose que l'énergie cinétique du mouvement de la particule est "indépendante de son environnement". Autrement dit, lorsqu'on démontre l'équipartition de l'énergie entre deux systèmes en interaction (ici : les molécules d'une part, la particule de l'autre), on suppose l'indépendance des probabilités d'état des deux systèmes (voir par exemple (7)), à la condition près de respecter les conditions imposées comme celle de la valeur de l'énergie totale. Mais cette hypothèse n'est pas satisfaite ici : il y a un certain entrainement du fluide par la particule et les paramètres qui échangent de l'énergie (vitesse de la particule et vitesse des molécules qui la heurtent) n'ont pas des valeurs indépendantes. On peut rétablir l'indépendance en n'attribuant comme énergie thermique aux molécules avoisinant la particule que celle qui correspond à la différence entre leurs vitesses réelles et les vitesses moyennes -hydrodynamiques) aux points coincidants. Simultanément il faut attribuer à la particule toute l'énergie cinétique hydrodynamique supplémentaire. La vitesse quadratique moyenne de la particule en est donc réduite d'autant.

Ces deux effets jouent en sens inverse sur le coefficient de diffusion, Nous allons montrer qu'il se compensent parfaitement sous certaines hypothèses (mais bien entendu la vitesse quadratique moyenne sera plus faible que celle qu'on utilise dans da démonstration de LANGEVIN.

FORME MODIFIÉE POUR L'ÉQUATION DU MOUVEMENT

Pour remplacer l'équation (3), nous devons déterminer quelle est la résistance hydrodynamique que le fluide oppose au mouvement de la particule à l'instant t, connaissant toute l'histoire de ce mouvement jusqu'à l'instant t. Or, avec les hypothèses de l'hydrodynamique, si l'on pose que le fluide a au contact de la particule la même vitesse que celle-ci et qu'il est au repos à l'infini, le mouvement du fluide est parfaitement défini, donc aussi la valeur de la résistance.

Les équations du mouvement du fluide sont l'équation de continuité (condition d'incompressibilité) et les équations dynamiques. Si dans ces dernières, on néglige le terme d'accélération, on aboutit finalement à la formule de Stokes. On sait que si on conserve ce terme d'accélération, on est en présence d'équations non linéaires. Dans ces conditions, on ne sait pas résoudre le problème posé ; d'ailleurs on ne peut plus traiter séparément les mouvements selon les trois directions, ni même le mouvement brownien de translation et celui de rotation. Nous opérerons la linéarisation habituelle qui consiste à assimiler la dérivée totale par rapport au temps à la dérivée partielle au point considéré comme fixe par rapport à la particule.

Divers auteurs depuis BOUSSINESQ (8) ont résolu le problème dans ces conditions et établi l'expression suivante pour la résistance :

$$(7) \qquad R(t) = 6\pi\eta r \frac{dx}{dt} + \frac{2}{3}\pi r^3 \rho \frac{d^2x}{dt^2} + 6r^2\sqrt{\pi\eta\rho}\int_{-\infty}^{t} \frac{\frac{d^2x}{d\tau^2}(\tau)\,d\tau}{\sqrt{t-\tau}}$$

où ρ est la masse spécifique du fluide.

On trouvera dans l'ouvrage de VILLAT (9) une étude détaillée de cette solution.

On déduit de là immédiatement l'équation du mouvement modifiée:

$$(8) \qquad \left(m + \frac{2}{3}\pi r^3 \rho\right)\frac{d^2x}{dt^2} + 6\pi\eta r \frac{dx}{dt} + 6r^2\sqrt{\pi\eta\rho}\int_{-\infty}^{t} \frac{\frac{d^2x}{d\tau^2}(\tau)\,d\tau}{\sqrt{t-\tau}} = F(t)$$

où F (t) est de nouveau l'écart aléatoire (de moyenne nulle) de la résultante des forces de pression. (1)

Notons ici les hypothèses faites (nous les discuterons <u>in fine</u>) :

Hypothèse I - Les équations des fluides continus donnent l'espérance mathématique des composantes réelles de la pression.

Hypothèse II - On peut négliger les termes quadratiques dans les équations du mouvement du fluide.

Hypothèse III - On peut assimiler la vitesse hydrodynamique du fluide au contact de la particule à la vitesse de la particule.

Hypothèse IV - On peut assimiler l'accélération hydrodynamique du fluide au contact de la particule à l'accélération de la particule (ce qui suppose une fluctuation pas trop rapide de cette dernière).

CONSIDÉRATIONS GÉNÉRALES SUR L'ÉQUATION DU MOUVEMENT

L'étude du mouvement brownien se ramène donc à l'inversion de l'équation (8). Si on se donne la fonction F (t), la vitesse de la particule $U(t) = \frac{dx}{dt}$ sera donnée par la solution de l'équation (8) (qui, on peut le montrer, existe et est unique). Puisque la fonction F (t) est aléatoire, il

---

(1) Tandis que le présent travail était en cours, nous avons découvert que l'idée d'étudier le remplacement de l'équation (3) par l'équation (8) se trouvait déjà suggérée dans une note en bas de page d'un article de WEYSSENHOFF [10] sur la résistance des fluides, note ainsi rédigée : "Es hätte sich vielleicht gelohnt, die verschiedenen Theorien der Brownschen Bewegung, bei welchen die Gültigkeit des Stokesschen Gesetzes sogar für die einzelnen Zickzackwege der unregelmäszigen Bahn vorausgesetzt wird, im Lichte der eben erwähnten theoretischen Untersuchungen durchzudiskutieren".

en sera de même de la fonction U (t), et les problèmes mathématiques se ramèneront à l'étude de la correspondance des propriétés stochastiques de ces deux fonctions. (1)

Cette étude rentre dans un cadre classique (voir par exemple (11) dont divers résultats seront utilisés par la suite). En effet on passe de F (t) à U (t) par l'intermédiaire d'un "filtre linéaire" et il suffira de caractériser ce dernier par son "gain" pour pouvoir répondre aux diverses questions qui se posent, Ce "gain" sera lié de façon simple à la "réponse percussionnelle".

Pratiquement la fonction aléatoire F (t) a des propriétés particulières. Tout d'abord elle est "stationnaire d'ordre deux", et il en sera par suite de même pour U (t). D'autre part, considérons les ordres de grandeur mis en jeu. F (t) est la superposition d'impulsions indépendantes dûes à chacune des molécules "heurtant" la particule ; pour des températures, des pressions et des natures de fluide courants, il y a plus de 10^{15} chocs par seconde et micron carré de surface, chaque "choc" durant 10^{-12} à 10^{-10} secondes. La réponse percussionnelle s'étend sur une durée qu'on peut caractériser par le paramètre $\frac{m}{6\pi\eta r}$ qui vaut pratiquement 10^{-6} à $10^{-5} r^2$ secondes (r en microns), pour des fluides non exceptionnellement visqueux. On pourra donc assimiler F (t), par rapport à U (t), à un "phénomène à corrélation microscopique" (ou "à spectre blanc") et U (t) sera quasi-laplacienne. Dans ces conditions, le problème est équivalent à celui de l'observation du "bruit pur" à travers un filtre linéaire. Notons enfin que le troisième ordre de grandeur mis en jeu, celui des intervalles entre les observations visuelles, est à son tour extrêmement grand devant les deux autres, de sorte que U (t) est à son tour à corrélation microscopique par rapport à des phénomènes à cette échelle et que x (t) peut être assimilé à une fonction aléatoire à accroissements indépendants ou "mouvement brownien" mathématique.

CALCUL DU GAIN

La façon la plus commode d'utiliser l'équation (8) consiste à déterminer préalablement le "gain" du filtre qu'elle représente. Le calcul revient à résoudre cette équation par la transformation de Fourier, ce qui a été déjà fait par divers auteurs (voir à ce sujet (6)),(10), Par défini-

---

(1) U (t) étant une fonction aléatoire, l'intégrale qui intervient dans l'équation (8) doit être considérée comme une intégrale stochastique. On peut montrer que celle-ci est définie en moyenne quadratique avec les précisions que nous indiquerons au sujet de F (t). On sait en effet qu'il suffit pour cela de prouver l'existence d'une limite finie lorsque $T \longrightarrow -\infty$ de l'intégrale double suivante ;

$$\int_T^t \int_T^t \frac{E\left[U'(\tau)\, U'(\tau')\right] d\tau\, d\tau'}{\sqrt{(t-\tau)(t-\tau')}}$$

c'est-à-dire

$$2\int_0^{-T} \frac{d^2\Gamma(u)}{du^2} \left[2L\left(\sqrt{-T-u}+\sqrt{-T}\right) - Lu \right] du$$

où Γ (u) est la fonction de correlation de U (t). Or les propriétés de Γ (u) assurent cette convergence.

tion, si à la fonction F (t) = $e^{2\pi i \nu t}$, l'équation (8) fait correspondre la fonction :

$$U(t) = G(\nu)\, e^{2\pi i \nu t}$$

G (ν) est le gain(complexe) relatif à la fréquence ν ;

En portant ces expressions de F et U dans l'équation (8), il vient :

$$G(\nu)\left[\left(m-\frac{2}{3}\pi r^3\rho\right) 2\pi i\nu + 6\pi\eta r + 6r^2\sqrt{\pi\eta\rho}\cdot 2\pi i\nu\int_{-\infty}^{-t}\frac{e^{-2\pi i\nu(t-\tau)}d\tau}{\sqrt{t-\tau}}\right] = 1$$

or l'intégrale écrite vaut :

$$\int_0^{+\infty}\frac{e^{-2\pi i\nu z}dz}{\sqrt{z}} = \frac{1}{\sqrt{2i\nu}}$$

quel que soit ν réel, en prenant pour détermination de la racine carrée celle qui est à partie réelle positive. D'où :

$$(9)\quad \frac{1}{G(\nu)} = \left(m+\frac{2}{3}\pi r^3\rho\right) 2\pi i\nu + 6\pi\eta r + 6r^2\pi\sqrt{\pi\eta\rho}\,\sqrt{2i\nu}$$

(avec la même détermination de la racine carrée).

COEFFICIENT DE DIFFUSION

Pour voir ce que devient la formule d'EINSTEIN, nous avons à calculer la limite de $\frac{E(x^2)}{t}$ lorsque t augmente indéfiniment. On a :

$$x = \int_0^t U(\tau)\, d\tau$$

et, d'après un calcul classique tenant compte de la stationnarité de U :

$$E(x^2) = t\int_{-t}^{+t}\left(1-\frac{|\tau|}{t}\right)\Gamma_U(\tau)\, d\tau$$

$\Gamma_U(u)$ étant la fonction de corrélation de U (t). D'où

$$\text{limite}\,\frac{E(x^2)}{t} = \int_{-\infty}^{+\infty}\Gamma_U(\tau)\, d\tau = f_U(o) = |G(o)|^2 f_F(o)$$

si $f_U(\nu)$ et $f_F(\nu)$ sont les densités de répartition spectrale du U et F, soit :

$$(10)\ \text{limite}\,\frac{E(x^2)}{t} = \frac{f_F(o)}{(6\pi\eta r)^2} = \frac{\int_{-\infty}^{+\infty}\Gamma_F(\tau)d\tau}{(6\pi\eta r)^2}$$

Si l'on admet, d'après l'analyse des phénomènes qui a été faite plus haut, que la fonction aléatoire F (t) est indépendante du mouvement de la particule, autrement dit qu'elle représente le même processus que les fluctuations de pression subies par une particule identique maintenue im-

mobile, il en résulte que f_F (0) et par suite le coefficient de diffusion sont indépendants de la masse spécifique de la particule. La démonstration de LANGEVIN, valable quand la masse spécifique du fluide est négligeable devant celle de la particule, aboutit donc à une forme (la formule d'Einstein) exacte dans le cas général (sous les hypothèses énoncées plus haut).

VITESSE QUADRATIQUE MOYENNE

La fonction de corrélation de la fonction aléatoire U (t) se déduit de façon simple de celle de F (t) et du gain G (ν). Ce qui nous intéresse spécialement au point de vue physique, c'est la valeur de E (U^2). On sait (11) que l'on a :

(11) $$E(U^2) = \int_{-\infty}^{+\infty} G(\nu) \cdot \overline{G(\nu)}\, f_F(\nu)\, d\nu$$

Mais nous avons dit que F (t) est à corrélation microscopique par rapport à U (t), autrement dit à spectre uniforme dans la zone des fréquences où le gain G (ν) n'est pas pratiquement nul. Par suite :

(12) $$E(U^2) = f_F(0) \int_{-\infty}^{+\infty} G(\nu)\, \overline{G(\nu)}\, d\nu$$

$\overline{G(\nu)}$ est la quantité imaginaire conjuguée de G (ν), en fait égale à G (−ν). A partir de (9), on calculera aisément l'intégrale indiquée, car il s'agit d'une fraction rationnelle en $\sqrt{\nu}$ (dénominateur du 4° degré dont on peut préciser les racines, puis les associer par valeurs conjuguées) et il suffira de faire le changement de variable $\sqrt{\nu} = z$ puis la décomposition en éléments simples.

On pourra d'autre part exprimer $f_F(\nu)$ en fonction de $\lim \frac{\overline{x^2}}{t}$ d'après (10). La relation ainsi obtenu entre $\overline{U^2}$ et $\frac{\overline{x^2}}{t}$ suppose uniquement de F (t) qu'elle est à corrélation microscopique. Le résultat de ces calculs est le suivant :

(13) $$\lim \frac{\overline{x^2}}{t} = \frac{m\overline{U^2}}{3\pi\eta r} \cdot f(\alpha)$$

$\left(\alpha = \frac{\rho}{\rho_0}\right.$, ρ_0 masse spécifique de la particule)

(14) $$f(\alpha) = \frac{\left(1+\frac{\alpha}{2}\right)\sin\varphi\cos\varphi}{2(1+\cos\varphi)\frac{\varphi}{\pi} - \sin\varphi} \quad \text{avec} \quad \cos\varphi = \frac{7\alpha-4}{2\alpha+4} \quad 0 \leqslant \varphi \leqslant \pi$$

pour $0 < \alpha < \frac{8}{5}$

(15) $$f(\alpha) = \frac{\left(1+\frac{\alpha}{2}\right)\operatorname{sh}\varphi\operatorname{ch}\varphi}{2(1+\operatorname{ch}\varphi)\frac{\varphi}{\pi} - \operatorname{sh}\varphi} \quad \text{avec} \quad \operatorname{ch}\varphi = \frac{7\alpha-4}{2\alpha+4} \quad \text{pour } \alpha > \frac{8}{5}$$

Voici quelques valeurs numériques de la fonction $f(\alpha)$

α	0	0,04	0,08	0,25	1	1,6	2	4	8
$f(\alpha)$	1	1,48	1,72	2,44	4,84	6,59	7,72	13,3	24,2

Si on admet en outre que la limite de $\frac{\overline{x^2}}{t}$ satisfait bien à la formule d'Einstein (hypothèse supplèmentaire sur $f_F(o)$), on voit aisément que l'on a :

$$m\,\overline{U^2} = \frac{RT}{N}\,\frac{1}{f(\alpha)} \tag{16}$$

Ainsi $f(\alpha)$ apparait comme le facteur dont est réduit l'énergie cinétique moyenne de translation selon une direction, par rapport à la valeur d'équipartition, si on tient compte de l'inertie du fluide. A titre de vérification, il est possible, en utilisant la solution de VILLAT (voir le paragraphe suivant) de calculer l'énergie hydrodynamique moyenne contenue dans le fluide ; on l'ajoutant à celle de la particule, on constate bien que l'on retrouve dans tous les cas la valeur d'équipartition.

RÉPONSE PERCUSSIONNELLE

On peut aussi déduire de l'équation (8) la réponse percussionnelle R (t) du filtre étudié (c'est-à-dire la fonction U (t) obtenue lorsque F (t) est une "fonction impulsion" ou "fonction de Dirac"). La connaissance de R (t), par laquelle on définit généralement les filtres, permet de déduire les diverses propriétés du filtre.

On peut obtenir R (t) à partir du gain par la formule:

$$R(t) = \int_{-\infty}^{+\infty} e^{2\pi i \nu t}\, G(\nu)\, d\nu \tag{17}$$

On peut aussi l'obtenir par résolution directe de l'équation (8). VILLAT (9) étudie en détail la solution de cette équation, lorsque F (t) est la "fonction saut" égale à 0 pour $t < 0$ et 1 pour $t > 0$). Cette solution semble dûe à BOGGIO (12) Par dérivation de cette solution on obtient la réponse percussionnelle, qu'on peut mettre sous la forme :

(18) U (t) = R (t) = 0 pour $t < 0$

$$= \frac{3}{\pi r^3 \sqrt{\pi(\rho + 2\rho_o)(5\rho - 8\rho_o)}} \left\{ \sqrt{\alpha}\; e^{\alpha P t} \int_{\sqrt{\alpha P t}}^{+\infty} e^{-u^2} du - \sqrt{\beta}\; e^{\beta P t} \int_{\sqrt{\beta P t}}^{+\infty} e^{-u^2} du \right.$$

pour $t > 0$

où α et β sont les racines réelles ou imaginaires de :

$$x^2 + \frac{4\rho_o - 7\rho}{2\rho_o + \rho}\, x + 1 = 0 \tag{19}$$

et où

$$P = \frac{6\pi\eta\, r}{\frac{4}{3}\pi r^3 \left(\rho_o + \frac{\rho}{2}\right)}$$

(et en adaptant la détermination de α et β à celle de la racine de $5\rho - 8\rho_o$)

A partir de R (t), on aurait pu obtenir directement la formule (13) en utilisant :

$$U^2 / \lim \frac{x^2}{t} = \frac{\int_0^{+\infty} R^2(t)\,dt}{\left[\int_0^{+\infty} R(t)\,dt\right]^2}$$

Le calcul explicite est effectivement possible.

Si au lieu de R^2, on porte dans l'intégrale du numérateur

$$R^2 + \frac{\rho}{\rho_0}\iiint V^2 dx\,dy\,dz$$

(intégrale étendue à tout le fluide du carré de la vitesse en chaque point du fluide), on peut vérifier que l'énergie totale satisfait au principe d'équipartition.

CRITIQUE DES HYPOTHÈSES

Il parait difficile d'apprécier l'écart que les hypothèses faites introduisent entre le calcul et la réalité. Mais nous allons montrer que cet écart est d'autant plus faible que la particule est plus grosse, de sorte que les formules que nous avons établies sont valables asymptotiquement pour un rayon suffisamment grand.

En effet la vitesse quadratique moyenne diminue comme la puissance $-\frac{1}{2}$ de la masse, c'est-à-dire la puissance $-\frac{3}{2}$ du rayon ; donc :

1) le gradient de vitesse et l'accélération du fluide au voisinage de la particule diminuent quand le rayon croit et les équations de l'hydrodynamique des fluides continus sont de plus en plus applicables au mouvement des molécules. (Voir par exemple à ce sujet (13)). En particulier les effets de "relaxation" dans le fluide s'atténuent.

2) le nombre de Reynolds du mouvement de la particule diminue comme la puissance $-\frac{1}{2}$ du rayon, de sorte que les termes quadratiques sont de plus en plus négligeables, la linéarisation de plus en plus valable.

3) l'importance du "glissement" du fluide sur la particule est de plus en plus négligeable, en ce qui concerne l'assimilation de la vitesse du fluide au voisinage de la particule à la vitesse de la particule.

4) Comme on l'a vu plus haut, le caractère microscopique de la corrélation de F (t) croit avec le rayon.

5) il reste le point le plus délicat et le plus important, à savoir l'hypothèse sur l'accélération. Il est certain que la formule (7) pour la résistance n'est pas valable pour une accélération trop fluctuante parce que les molécules qui heurtent la particule viennent d'une zone où les variations d'accélération sont ressenties déjà avec un certain retard et atténuées. Or l'accélération U' (t) fluctue de façon analogue à F (t), soit très rapidement et même - on peut le voir aisément - de plus en plus quand r croit. Quelle répercussion ce fait a-t-il sur les résultats obtenus ? On peut faire l'hypothèse plausible que l'équation à substituer à (7) et où interviendraient

les grandeurs moléculaires telles que le libre parcours moyen et le temps de relaxation moléculaire, serait encore linéaire. Par suite l'étude faite serait encore valable à condition de modifier l'expression du gain $G(\nu)$. Mais l'expression de ce gain resterait la même pour ν assez petit. L'intégrale (12) serait modifiées de ce fait. Or $\frac{G(\nu)}{G(o)}$ décroit comme $\frac{1}{r^2}$ de sorte que la valeur de $|\nu|$ à partir de laquelle la quantité $G(\nu)\overline{G(\nu)}$ a une contribution négligeable dans l'intégrale (12) décroit comme $\frac{1}{r^2}$. Donc à ce point de vue encore les conditions pour la validité des formules sont d'autant mieux remplies que r est plus grand.

LA DÉMONSTRATION D'EINSTEIN

Il est intéressant d'examiner la démonstration initiale d'Einstein (voir (1) et (2)) à la lumière des considérations qui précèdent. Rappelons qu'Einstein introduit un champ de forces uniforme auxiliaire (n'affectant que les particules et non le fluide) et considère un système en équilibre, comprenant un grand nombre de particules dans un fluide, mais dans un volume assez grand pour qu'il n'y ait pas d'interaction entre les particules (ou mieux encore on peut considérer une probabilité de présence d'une particule unique).

Le théorème de BOLTZMANN (14) donne la loi de répartition spatiale des particules (loi de probabilité) par la considération de leur énergie potentielle ; il suffit alors d'écrire que les mouvements de diffusion compensent le déplacement d'ensemble sous l'influence du champ de forces, pour déterminer le coefficient de diffusion brownienne, qui se relie aisément à $E(x^2)/t$. (On évite ainsi de passer par la détermination de $E(U^2)$).

Cette démonstration échappe aux deux objections physiques que nous avons retenues. D'autre part l'énergie potentielle des particules est indépendante de l'énergie cinétique des molécules qui la heurtent et le théorême de Boltzmann doit être valable. D'autre part on peut appliquer la loi de Stokes pour évaluer le déplacement uniforme dû au champ.

Cependant il faut remarquer que le raisonnement suppose l'indépendance du mouvement dû au champ et du mouvement brownien, ce qui revient à admettre implicitement la validité de la linéarisation des équations hydrodynamiques. Telle est donc finalement la limitation de validité de la formule d'Einstein.

CONCLUSION

Nous avons étudié l'effet de l'inertie du fluide sur le mouvement brownien d'une particule qui se trouve plongée dans ce fluide et nous avons montré, moyennant certaines hypothèses.

1°) que la formule d'Einstein reste valable (comme le laissait supposer la démonstration initiale d'Einstein).

2°) que la vitesse quadratique moyenne de la particule est en fait plus petite que celle qu'on calcule par l'application du principe d'équipartition de l'énergie. Le coefficient de réduction ne dépend que du rapport des masses spécifiques du fluide et de la particule et croit avec ce rapport.

3°) que les hypothèses formulées (et par suite les conclusions obtenues) sont d'autant plus valables que la particule est plus grosse, (Evidemment, l'autre cas limite est celui du mouvement brownien d'une molécule de fluide, qui, en particulier, satisfait au principe d'équipartition).

L'intérêt pratique de ces résultats est limité par l'inaccessibilité expérimentale actuelle de la vitesse quadratique moyenne (l'élargissement par effet Doppler-Fizeau d'une raie lumineuse émise ou réfléchie par une particule en mouvement brownien est trop faible pour être mesurable). Il reste cependant à étudier la valeur du coefficient de diffusion pour les petits rayons (en dessous du dixième de micron).

L'intérêt théorique nous parait résider dans la nécessité qui se dégage de ces résultats d'établir dans quelles conditions exactes s'applique le principe d'équipartition de l'énergie. En effet les coordonnées et les composantes de la vitesse des molécules et de la particule appartiennent bien à une représentation canonique du système physique considéré et pourtant le principe d'équipartition ne s'applique pas. Ce fait remet en cause le choix de la densité de probabilité dans l'espace d'extension en phase.

BIBLIOGRAPHIE

1 A. EINSTEIN, Ann. Der Phys. 17, 549 (1905)

2 A. EINSTEIN, Ann. Der Phys. 19, 371 (1906)

3 SMOLUCHOWSKI, Bull. Acad. Cracovie (1906), 577

4 A. EINSTEIN, Zeit. f. Elektrochemie 13, 41 (1907)

5 LANGEVIN, C.R. Ac. Sc., 146, 530 (1908)

6 TCHEN CHAN-MOU, Mean value and correlation problems connected with the motion of small particles suspended in a turbulent fluid. The Hague, Martinus Nijhoff (1947)

7 E. BOREL, Mécanique statistique classique - -Gauthier-Villars (1925)

8 BOUSSINESQ, C.R. Ac. Sc. p, 935 (1885)

9 H. VILLAT, Leçons sur les fluides visqueux. Gauthier-Villars (1943)

10 J. WEYSSENHOFF, Betrachtungen über den Gültigkeitsbereich der Stokesschen und der Stokes-Cunninghamschen Formel, Ann. Der Phys. 62, 1 (1920)

11 FORTET ET BLANC-LAPIERRE, Théorie des fonctions aléatoire, Masson (1953)

12 T. BOGGIO, Rend. Acc. Lincei 16, p. 613 et 730 (1907)

13 ROCARD, L'hydrodynamique et la théorie cinétique des gaz. Gauthier-Villars (1932)

14 BOLTZMANN, Leçons sur la théorie des gaz I et II. Gauthier-Villars (1902 et 1905)

Les références (1), (2) et (4) sont réunies (traduites en anglais) dans :

A. EINSTEIN, Investigations on the theory of the brownian movement : London Methuen and C°, 1926.

J. & R. SENNAC, Imprimeurs
54, Fg Montmartre, PARIS (9e)